# Bootstrap inference in functional linear regression models

by

**Hyemin Yeon**

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Daniel John Nordman, Co-major Professor
Xiongtao Dai, Co-major Professor
Mark Steven Kaiser
Yumou Qiu
Lily Wang



Iowa State University

Ames, Iowa

2023

## DEDICATION

To my loving family

# TABLE OF CONTENTS

# LIST OF TABLES

**Page**

# LIST OF FIGURES

# ACKNOWLEDGMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this thesis. First and foremost, Dr. Daniel Nordman and Dr. Xiongtao Dai for their guidance, patience and support throughout this research and the writing of this thesis. Their insights and words of encouragement have often inspired me and renewed my hopes for completing my graduate education. I would also like to thank my committee members for their efforts and contributions to this work: Dr. Mark Kaiser, Dr. Yumou Qiu, and Dr. Lily Wang. I would additionally like to thank Dr. Sara Lopez-Pintado at Northwestern University for her guidance and support of the work on data depth.

In addition to the faculty members at Iowa State University, I also wish to thank some academic people from Seoul National University (SNU). I thank Dr. Johan Lim, my major professor in the master program in Statistics at SNU. His kind words have encouraged me to pursue and complete this doctoral program. I thank Dr. Joonpyo Kim, who was my best academic colleague at SNU. His help of completing the coursework in my master program made me much solid in theoretical statistics, which has been the driving force behind completing my research. I thank Dr. Hanbaek Lyu, who is my best academic friend from SNU. The conversation with him and his practical advice have given relief whenever I was concerned about my future career.

Finally, I want to give thanks to my family including my wife, Sooran Kim, and my two kids. They have been the source of happiness in my life, especially in Ames. I am also thankful to my parents and big brother in South Korea, who always supports me unquestioningly.

## ABSTRACT

We consider functional linear regression models (FLRMs) with functional regressor and scalar response, where the inference of the slope function is an important problem. However, even though asymptotic inference methods exist in FLRMs, these methods are limited in applicability because a wrong scaling factor is used; truncation bias in the limit is neglected; or only homoscedastic errors are assumed, which may not happen in practice. Consequently, it is necessary to develop alternative inference methods, such as bootstrap, that use the correct scaling, accommodate possible bias, and are valid even under heteroscedasticity. In this thesis, we introduce three bootstrap methods in FLRMs, namely the residual bootstrap, paired bootstrap, and wild bootstrap. Their theoretical validities are established, and their performances are numerically demonstrated. Central limit theorems for the projection are studied as well, which are fundamental results themselves and are basis to verify bootstrap validity.

# CHAPTER 1. GENERAL INTRODUCTION

This dissertation investigates three bootstrap methods for inference in functional linear regression models (FLRMs) with functional regressor and scalar response. Even though asymptotic inference methods exist in FLRMs, these methods are limited to in applicability because a wrong scaling factor is used; truncation bias in the limit is neglected; or only homoscedastic errors are assumed, which may not happen in practice. Consequently, it is necessary to develop alternative inference methods, such as bootstrap, that use the correct scaling, accommodate possible bias, and are valid even under heteroscedasticity.

The following three research papers constitute the dissertation:

Paper 1: Bootstrap inference in functional linear regression models with scalar response.

Paper 2: Bootstrap inference in functional linear regression models with scalar response under heteroscedasticity.

Paper 3: An initial theoretical work on wild bootstrap for functional linear regression.

The first two papers treat two different bootstrap methods in FLRMs under either homoscedastic or heteroscedastic error assumptions. In either case, a suitable central limit theorem (CLT) justifies the developed bootstrap methods. The last paper contains initial theoretical justification for a wild bootstrap method in FLRMs.

Paper 1 considers a new residual bootstrap method in FLRMs under homoscedasticity. The proposed residual bootstrap is theoretically shown to be consistent and is widely applicable for constructing both confidence and prediction regions at target regressor points. The method is also extendable to simultaneous regions, which is less tractable by normal approximation. The establishment of the bootstrap further involves generalizing, refining, and correcting a foundational CLT for functional linear regression.

Paper 2 develops a new paired bootstrap method in FLRM under heteroscedasticity. A novel CLT is established under heteroscedasticity; CLTs have not even been investigated in this case. The paper then shows the proposed paired bootstrap provides valid inference in FLRMs under heteroscedastic error assumptions, while it also exhibits good numerical performance in homoscedastic cases. Interestingly, the paired bootstrap is also shown to fail if this is implemented in a naïve way. As an application of the paired bootstrap, a novel hypothesis test for projections is developed, which are supported both theoretically and numerically.

Paper 3 focuses on multiplier wild bootstrap as an alternative to paired bootstrap in FLRMs under heteroscedasticity, particularly for large data cases. Since paired bootstrap repeats computing (pseudo-)inverse covariance operators in every re-sample, wild bootstrap that uses just residuals is beneficial in terms of computing speed. This paper provides wild bootstrap consistency and its theoretical details.

# CHAPTER 2. BOOTSTRAP INFERENCE IN FUNCTIONAL LINEAR REGRESSION MODELS WITH SCALAR RESPONSE

Modified from a paper accepted by *the Bernoulli Journal*

Hyemin Yeon[1], Xiongtao Dai[2], and Daniel John Nordman[1]

[1]Department of Statistics, Iowa State University

[2]Division of Biostatistics, University of California, Berkeley

**Abstract**

In fitting linear regression models for functional data, a complicating factor with regressors as random curves is that regression estimators have complex distributions, due to issues in bias and scaling. Bias arises because the target slope function is infinite-dimensional, while finite-sample estimators necessarily involve truncations. To approximate sampling distributions, we develop a residual bootstrap method. Despite the parametric regression problem, the bootstrap for functional data requires a development that resembles resampling for nonparametric regression with multivariate regressors. Essentially, original- and bootstrap-data estimators require coordination in the truncation levels to remove bias (akin to tuning parameter choices). The resulting bootstrap has wide applicability for constructing both confidence and prediction regions at target regressor points, and with coverage properties even holding conditionally on data regressors; the method also extends to simultaneous regions. Establishment of the bootstrap further involves generalizing, refining, and correcting a foundational central limit theorem for functional linear regression. Numerical studies verify our theory, showing that the bootstrap performs better than normal approximations, and also suggest a rule of thumb for setting the truncation levels. The bootstrap method is illustrated with an application to wheat spectrum data.

## 2.1 Introduction

Functional data analysis (FDA) has seen intensive development during the last two decades to address fundamental data units being trajectories, surfaces, and more general functions (cf. [32]). Overviews of FDA may be found in several reference textbooks, such as [12, 20, 21, 25, 29]. Our work focuses on the functional linear regression model (FLRM), a generalization of the classical linear regression model to the case where the predictor is a function. FLRM and its extensions are highly relevant in practice and have been applied in subject areas such as bio-medicine [15, 30] and agronomy [31], among others; see, also, [27] for a review of the applications.

The FLRM with scalar response can be written as

$$Y = \langle \beta, X \rangle + \varepsilon, \tag{2.1}$$

where $\beta$ is a slope function and $X$ is a random function, both taking values in a Hilbert space $\mathbb{H}$ with inner product $\langle \cdot, \cdot \rangle$; $Y$ represents a scalar response; and $\varepsilon$ is a random error with mean zero and finite variance, that is uncorrelated with $X$. For example, a random function $X$ is commonly modeled in $\mathbb{H} = L^2([0,1]) \equiv \{f : [0,1] \to \mathbb{R} \mid \int_0^1 f^2(t)dt < \infty\}$, the space of all square-integrable functions supported on $[0,1]$, equipped with inner product $\langle f_1, f_2 \rangle \equiv \int_0^1 f_1(t) f_2(t) dt$ for $f_1, f_2 \in \mathbb{H}$.

Based on a random sample $\{(Y_i, X_i)\}_{i=1}^n$ of size $n$ from the model (2.1), estimation of the slope function $\beta \in \mathbb{H}$ is challenging due to an ill-posed problem with inversion of the sample covariance operator of $\{X_i\}_{i=1}^n$. Consequently, an estimator $\hat{\beta}_{h_n}$ of $\beta$ is commonly constructed from the functional principal components (FPCs) of this covariance operator (cf. [6, 19]). This involves the selection of a number $h_n$ of FPCs where the corresponding eigenspace determines a finite-dimensional approximation to the slope function $\beta$. The latter, though, is typically infinite-dimensional and thus bias necessarily occurs in the approximation, analogous to the nonparametric regression setting [18]. While the regression estimator $\hat{\beta}_{h_n}$ is consistent under smoothness assumptions [19], Cardot, Mas, and Sarda [7] (herafter referred to as [CMS]) critically showed that $a_n(\hat{\beta}_{h_n} - \beta)$ cannot converge in distribution to any non-degenerate random function taking values in $\mathbb{H}$, under any scaling sequence $\{a_n\}$ such that $a_n \to \infty$ as $n \to \infty$. However, importantly for inference, [CMS] also showed that the projections defined as $\sqrt{n/h_n}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ can have distributional limits and satisfy a type of the Central Limit Theorem (CLT), for $X_0$ denoting a random regressor point that is independent of the sample. We focus attention on such projections, noting though that the CLT involved actually requires a bit more development. In particular, we consider the bootstrap as a device for approximating the sampling distributions of estimated projections $\langle\hat{\beta}_{h_n}, X_0\rangle$ from FLRMs.

Bootstrap methods for regression models, particularly the residual bootstrap, have a long history beginning with [14] and continuing to recent big-data regression problems (cf. [10]). However, these cases involve finite-dimensional regression parameters, which differs from FLRMs where the bootstrap must mimic estimation of infinite-dimensional slope function $\beta$ and accommodate the possible bias in this. Due to such difficulties, bootstrap methods for FLRMs have not received much development outside of important works by [16, 24]. [16] established a residual bootstrap for FLRMs in a specialized context where the target parameter was not the projection $\langle\beta, X_0\rangle$, but rather a biased version of this (c.f. $\langle\Pi_{h_n}\beta, X_0\rangle$ in Section 2.2). [24] investigated properties of percentile bootstrap confidence intervals for $\langle\beta, X_0\rangle$ under a modified residual bootstrap procedure. Essentially, bootstrap consistency in [16, 24] are not shown to hold conditionally on the regressors because of the dependency on the unconditional CLT developed in [CMS]. Neither of these previous works considered prediction intervals or simultaneous inference with bootstrap in FLRMs.

These aspects motivate us to study the residual bootstrap under a more general framework for FLRM. Our new contributions include accounting for possible bias and treating wider inference scenarios: calibrating either confidence regions for $\langle\beta, X_0\rangle$ or prediction regions for a new response $Y_0$, whether conditionally or unconditionally on a regressor $X_0$, and for both pointwise or simulatenous inference cases. The bootstrap approximations also capture the distribution of estimators (e.g., $\langle\hat{\beta}_{h_n}, X_0\rangle$), *conditionally* on the given regressors $\{X_i\}_{i=1}^n$, which goes beyond unconditional distributions considered by the CLTs from either [CMS] or [23], or the initial bootstrap works of either [16] or [24]. Our numerical studies also suggest that the bootstrap generally performs better than normal approximations and extends well to simultaneous intervals, where normal approximations become less tractable.

While our development is heavily influenced by the CLT work of [CMS] for projections from FLRMs, we refine and generalize those CLT results as another contribution. For a random new regressor $X_0$, the CLT there suggests $\sqrt{n/h_n}(\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle)$ has a normal limit, where $h_n$ denotes a number of FPCs. However, the scaling $\sqrt{n/h_n}$ is not generally valid and should be replaced by another scaling factor $\sqrt{n/t_{h_n}(X_0)}$ depending on $h_n$, the target regressor $X_0$, as well as the eigenvalues of the covariance operators in (2.1). The generalized CLT for FLRMs improves upon [CMS] by having a unified scaling for $X_0$ being random or given, and by further holding conditionally on any data regressors $\{X_i\}_{i=1}^n$ rather than only unconditionally. These findings form the basis for developing the new bootstrap results in FLRMs.

The organization of the paper is as follows. Section 2.2 outlines background on the FLRM (2.1) as well as the regressor estimator $\hat{\beta}_{h_n}$ based on FPCs and regularization. A generalized CLT result is presented in Section 2.3. Section 2.4 then describes the residual bootstrap method for FLRMs and establishes its validity for both prediction and estimation. Numerical studies appear in Section 2.5, while Section 2.6 provides a data application to illustrate the bootstrap method. Proofs are outlined in an appendix and further included in the supplement [34]. An R package is provided to find confidence and prediction intervals for FLRM projections based on either CLT or residual bootstrap.

## 2.2 Background on estimation for FLRMs

### 2.2.1 Model and identifiability

We suppose that the underlying Hilbert space $\mathbb{H}$ is separable throughout the paper and define the tensor product $x \otimes y : \mathbb{H} \to \mathbb{H}$ between two elements $x, y \in \mathbb{H}$ as a bounded linear operator $z \mapsto (x \otimes y)(z) = \langle z, x\rangle y$, for $z \in \mathbb{H}$. Without loss of generality, suppose that the regressor $X$ in the FLRM (2.1) has a finite second moment $\mathsf{E}[\|X\|^2] \equiv \mathsf{E}[\langle X, X\rangle] < \infty$ and zero mean $\mathsf{E}[X] = 0$, which is common in theory development for FLRM for ease of exposition (cf. [CSM]); in general, the responses and regressors in the data $\{(Y_i, X_i)\}_{i=1}^n$ can be centered by their respective sample means without affecting the results to follow. Write $\Gamma \equiv \mathsf{E}[X \otimes X]$ and $\Delta \equiv \mathsf{E}[YX]$ to respectively denote the covariance operator of $X$ and the cross-covariance of $X$ and $Y$, respectively. Then, we have the following functional version of the normal equation as

$$\Delta = \Gamma\beta. \tag{2.2}$$

This equation will be solved to identify the parameter $\beta \in \mathbb{H}$.

Let $T$ be a bounded linear operator on $\mathbb{H}$. The adjoint of $T$, denoted by $T^*$, is defined by $\langle Tx, y\rangle = \langle x, T^*y\rangle$ for $x, y \in \mathbb{H}$, and $T$ is said to be self-adjoint if $T = T^*$. A non-negative definite operator is a self-adjoint operator $T$ with the property that $\langle Tx, x\rangle \geq 0$ for each $x \in \mathbb{H}$, which also admits a square-root operator $T^{1/2}$ such that $(T^{1/2})^2 = T^{1/2}T^{1/2} = T$. If, for any bounded sequence $\{x_n\} \subseteq \mathbb{H}$, $\{Tx_n\}$ has a convergent subsequence in $\mathbb{H}$, then $T$ is said to be compact. Finally, if $\sum_{j=1}^{\infty} \|T\phi_j\|^2 < \infty$ holds for a complete orthonormal basis $\{\phi_j\}$ for $\mathbb{H}$, then $T$ is called a Hilbert–Schmidt operator.

With this background, the covariance operator $\Gamma$ is self-adjoint, non-negative definite, and Hilbert-Schmidt, and hence compact [21]. By spectral decomposition for compact self-adjoint operators, $\Gamma$ admits the decomposition

$$\Gamma = \sum_{j=1}^{\infty} \lambda_j (e_j \otimes e_j),$$

where $\lambda_j$ and $e_j$ respectively denotes the $j$th eigenvalue and the corresponding eigenfunction in $\mathbb{H}$ for $j \geq 1$. Here, $\{e_j\}$ forms an orthonormal system of $\mathbb{H}$, and $\{\lambda_j\}$ is a positive non-increasing sequence with $\lambda_j \to 0$ as $j \to \infty$. For the identifiability of $\beta$ in (2.2), we assume that $\ker\Gamma = \{0\}$ for simplicity of presentation as in previous works [7, 16]. The slope function is then written as

$$\beta = \Gamma^{-1}\Delta,$$

where $\Gamma^{-1} \equiv \sum_{j=1}^{\infty} \lambda_j^{-1}\pi_j$ and $\pi_j \equiv e_j \otimes e_j$, $j \geq 1$.

For later development, we define here some additional quantities related to $\Gamma \equiv \mathsf{E}[X \otimes X]$. For $h = 1, 2, \ldots,$ let

$$\Pi_h = \sum_{j=1}^{h} \pi_j, \tag{2.3}$$

be the projection onto the first $h$ eigenfunctions $\{e_j\}_{i=1}^{n}$ of $\Gamma$. Finally, for reference, the Karhunen–Loève expansion of $X$ is written as

$$X = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \xi_j e_j, \tag{2.4}$$

where the FPCs $\{\xi_j\}$ form a sequence of uncorrelated random variables with zero mean and unit variance (cf. [21]).

### 2.2.2 Regression estimator and regularization

For estimating the slope $\beta \in \mathbb{H}$, we consider a random sample $\{(Y_i, X_i)\}_{i=1}^{n}$ of $n$ paired observations from the model (2.1), namely,

$$Y_i = \langle \beta, X_i \rangle + \varepsilon_i, \quad i = 1, \ldots, n. \tag{2.5}$$

Regarding the distribution of $(\varepsilon_1, X_1)$, we assume only that $\mathsf{E}[\varepsilon_1 | X_1] = 0$ and $\mathsf{E}[\varepsilon_1^2 | X_1] \equiv \sigma_\varepsilon^2 \in (0, \infty)$ along with an accompanying property: for any integer $u \geq 0$, it holds almost surely that $\mathsf{E}[\varepsilon_1^2 \mathbb{I}(|\varepsilon_1| \geq u) | X_1] \leq f(u)$ for a non-random function $f(u)$ where $\lim_{u \to \infty} f(u) = 0$ (where $\mathbb{I}(\cdot)$ denotes the indicator function). Along with the conditional mean and variance, this integrability condition with conditional second moment is mild and holds trivially when errors $\varepsilon_i$ and regressors $X_i$ are independent as an important special case.

As counterparts to the covariances $\Gamma$ and $\Delta$ in the population normal equation (2.2), the sample versions are defined as $\hat{\Gamma}_n \equiv n^{-1} \sum_{i=1}^{n} (X_i \otimes X_i)$ and $\hat{\Delta}_n \equiv n^{-1} \sum_{i=1}^{n} Y_i X_i$. Here, $\hat{\Gamma}_n$ admits spectral decomposition $\hat{\Gamma}_n = \sum_{j=1}^{n} \hat{\lambda}_j \hat{\pi}_j$ where $\hat{\pi}_j \equiv (\hat{e}_j \otimes \hat{e}_j)$, in terms of the $j$th sample eigenvalue $\hat{\lambda}_j \geq 0$ and corresponding eigenvector $\hat{e}_j \in \mathbb{H}$. Inverting $\hat{\Gamma}_n$ is ill-posed because of the finite-sample nature and decaying eigenvalues, which complicates a sample analog of the parameter $\beta = \Gamma^{-1} \Delta$. To handle this issue (cf. [5, 7, 19, 16]), a regression estimator $\hat{\beta}_{h_n}$ of $\beta$ is defined by regularizing the inversion of $\hat{\Gamma}_n$, obtaining

$$\hat{\beta}_{h_n} = \hat{\Gamma}^{\dagger}_{h_n} \hat{\Delta}_n, \tag{2.6}$$

where $\hat{\Gamma}^{\dagger}_{h_n} \equiv \sum_{j=1}^{h_n} \hat{\lambda}_j^{-1} \hat{\pi}_j$ denotes a finite-sample approximation of $\Gamma^{-1} \equiv \sum_{j=1}^{\infty} \lambda_j^{-1} \pi_j$ based on a choice $h_n = 1, 2, \ldots$ of truncation level. That is, $h_n$ represents the number of eigenpairs from $\hat{\Gamma}_n$ used in estimation and may depend on the sample size $n$.

### 2.3 CLT for FLRM projections

We now describe some large-sample distributional properties of estimated projections $\langle \hat{\beta}_{h_n}, X_0 \rangle$ for later developing bootstrap inference in Section 2.4. As mentioned in Section 2.1, CLT results can hold for projections $\langle \hat{\beta}_{h_n}, X_0 \rangle$ formed with a new regressor point $X_0 \in \mathbb{H}$, where $X_0$ may be either random or fixed. A complication, though, is that this CLT holds most readily with biased centering $\langle \Pi_{h_n} \beta, X_0 \rangle$ rather than a target of $\langle \beta, X_0 \rangle$; here $\Pi_{h_n} \beta = \sum_{j=1}^{h_n} \langle \beta, e_j \rangle e_j$ is a truncated version of the parameter $\beta \equiv \sum_{j=1}^{\infty} \langle \beta, e_j \rangle e_j$, where $\Pi_{h_n}$ from (2.3) is the projection on the first $h_n$ eigenfunctions $\{e_j\}_{j=1}^{h_n}$ of $\Gamma$. A bias occurs due to the regularization step (2.6) in $\hat{\beta}_{h_n}$. A further complication is that, even with biased centering $\langle \Pi_{h_n} \beta, X_0 \rangle$, the CLT for $\langle \hat{\beta}_{h_n}, X_0 \rangle$ requires some refinement from the original work of [CMS], as given next.

Let $X_0$ be a new regressor observation under the model (2.1), independent of $\{(Y_i, X_i)\}_{i=1}^{n}$. [CMS] considered a CLT for the projection $[\langle \hat{\beta}_{h_n}, X_0 \rangle - \langle \Pi_{h_n} \beta, X_0 \rangle]$ with biased centering under mild assumptions that we also adopt. These conditions are

(A1) $\ker \Gamma = \{0\}$;

(A2) $\sup_{j \in \mathbb{N}} \mathsf{E}[\xi_j^4] < \infty$;

(A3) $\sum_{j=1}^{\infty} |\langle \beta, e_j \rangle| < \infty$;

(A4) $\lambda_j = \varphi(j)$ holds, at least with large $j$, for a convex positive function $\varphi : [1, \infty) \to \mathbb{R}$;

(A5) $\sup_{j \geq 1} \lambda_j j \log j < \infty$; and

(A6) $n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}\to 0$ as $n\to\infty$ for the sequence of eigengaps defined as $\delta_1\equiv\lambda_1-\lambda_2$ and $\delta_j\equiv\min\{\lambda_j-\lambda_{j+1},\lambda_{j-1}-\lambda_j\}$ for $j\geq 2$.

In brief, Condition (A1) ensures that $\beta$ is the unique solution to the normal equations (2.2). Condition (A2) implies a finite fourth moment $\mathsf{E}[\|X\|^4]<\infty$ for the regressor. Condition (A3) embodies a degree of smoothness for $\beta$. Both this and the convexity of $\{\lambda_j\}$ in (A4) may potentially be relaxed, but are imposed for simplicity. Condition (A5) is a decay condition on eigenvalues, while (A6) prescribes a decay rate on eigengaps $\{\delta_j\}$ in relation to the truncation level $h_n$ defining the estimator $\hat{\beta}_{h_n}$; these conditions are applied in proofs involving perturbation theory for functional data.

For a CLT with the FLRM projection $[\langle\hat{\beta}_{h_n},X_0\rangle-\langle\Pi_{h_n}\beta,X_0\rangle]$, or the counterpart $[\langle\hat{\beta}_{h_n},X_0\rangle-\langle\beta,X_0\rangle]$ with true parameter centering $\langle\beta,X_0\rangle$, we use a scaling factor given by $\{n/t_{h_n}(X_0)\}^{-1/2}$, involving a term

$$t_{h_n}(x)\equiv\sum_{j=1}^{h_n}\lambda_j^{-1}\langle x,e_j\rangle^2=\|(\Gamma_{h_n}^{\dagger})^{1/2}x\|^2,\quad x\in\mathbb{H}.\tag{2.7}$$

Due to regularization in defining the estimator $\hat{\beta}_{h_n}$ (cf. Section 2.2.2), the quantity (2.7) represents a norm involving a truncated version $\Gamma_h^{\dagger}\equiv\sum_{j=1}^{h_n}\lambda_j^{-1}(e_j\otimes e_j)$ of $\Gamma^{-1}=\sum_{j=1}^{\infty}\lambda_j^{-1}(e_j\otimes e_j)$. A sample analog of (2.7) is given as

$$\hat{t}_{h_n}(x)\equiv\sum_{j=1}^{h_n}\hat{\lambda}_j^{-1}\langle x,\hat{e}_j\rangle^2=\|(\hat{\Gamma}_{h_n}^{\dagger})^{1/2}x\|^2,\quad x\in\mathbb{H},\tag{2.8}$$

based on sample quantities and the finite-sample approximation $\hat{\Gamma}_{h_n}^{\dagger}\equiv\sum_{j=1}^{h_n}\hat{\lambda}_j^{-1}(\hat{e}_j\otimes\hat{e}_j)$ of $\Gamma^{-1}$ for defining $\hat{\beta}_{h_n}$ in (2.6).

With this scaling, our Theorem 1 next states a generalized CLT for FLRM projections under essentially the same weak conditions intended by [CMS]. Write $\mathcal{X}_n=\{X_1,\dots,X_n\}$ as the set of observed regressors and let $X_0$ again denote an independent regressor point under the model. In the following, let $\tilde{\mathsf{P}}(\cdot)$ denote either the conditional probability $\mathsf{P}(\cdot|\mathcal{X}_n,X_0)$ or $\mathsf{P}(\cdot|\mathcal{X}_n)$, where $X_0$ may be considered conditionally or unconditionally.

**Theorem 1** (Generalized/refined CLT for projections). *Under the FLRM (2.5), suppose that Conditions (A1)-(A6) hold along with* $h_n^{-1} + n^{-1/2} h_n^{5/2} (\log h_n)^2 \to 0$. *In addition, suppose* $h_n t_{h_n}(X_0)^{-1} = O_{\mathsf{P}}(1)$. *Then, as* $n \to \infty$, *we have*

*(i)*

$$\sup_{y \in \mathbb{R}} \left| \tilde{\mathsf{P}} \left( \sqrt{\frac{n}{t_{h_n}(X_0)}} \left[ \langle \hat{\beta}_{h_n}, X_0 \rangle - \langle \Pi_{h_n} \beta, X_0 \rangle \right] \leq y \right) - \Phi(y/\sigma_\varepsilon) \right| \xrightarrow{\mathsf{P}} 0,$$

*where* $\Phi$ *denotes the standard normal distribution function.*

*(ii)* $t_{h_n}(X_0)$ *and* $\hat{t}_{h_n}(X_0)$ *are equivalent in that, for any* $\eta > 0$,

$$\tilde{\mathsf{P}} \left( \left| \frac{\hat{t}_{h_n}(X_0)}{t_{h_n}(X_0)} - 1 \right| > \eta \right) \xrightarrow{\mathsf{P}} 0.$$

*Thus, the result in (i) also holds replacing* $t_{h_n}(X_0)$ *by the sample version* $\hat{t}_{h_n}(X_0)$.

Let $\check{\mathsf{P}}$ denote either the probability $\mathsf{P}(\cdot|X_0)$ or $\mathsf{P}$, where the observed regressors $\mathcal{X}_n$ are considered unconditionally. The unconditional CLT is then stated as below.

**Corollary 1.** *Under the assumptions of Theorem 1, as* $n \to \infty$, *we have*

$$\sup_{y \in \mathbb{R}} \left| \check{\mathsf{P}} \left( \sqrt{\frac{n}{t_{h_n}(X_0)}} \left[ \langle \hat{\beta}_{h_n}, X_0 \rangle - \langle \Pi_{h_n} \beta, X_0 \rangle \right] \leq y \right) - \Phi(y/\sigma_\varepsilon) \right| \xrightarrow{\mathsf{P}} 0,$$

*where the convergence above remains valid with estimated scaling* $\hat{t}_{h_n}$ *from (2.8).*

*Remark* 1. Theorem 1 again involves biased centering $\langle \Pi_{h_n} \beta, X_0 \rangle$. An analogous CLT holds with unbiased centering $\langle \beta, X_0 \rangle$, but requires more assumptions such as those required by Theorem 3; see Theorem 5 in the supplement. In this case, a specific example for the valid choice of the tuning parameter $h_n$ is given in Corollary 2 in Section 2.4.2.

Theorem 1 involves a mild condition that $h_n t_{h_n}(X_0)^{-1} = O_{\mathsf{P}}(1)$, so the scalings $t_{h_n}(X_0)$ does not have to scale as $h_n$. To give some examples, we provide some sufficient conditions on the sequence $\{\xi_j\}$ of the FPC scores for $h_n t_{h_n}(X_0)^{-1} = O_{\mathsf{P}}(1)$ to hold: (i) if $\mathsf{P}(\tau_1 \leq |\xi_j| \leq \tau_2) = 1$, where $0 < \tau_1 < \tau_2 < \infty$ for all integer $j \geq 1$, or (ii) if the average $h_n^{-1} \sum_{j=1}^{h_n} \xi_j^2$ converges to $V$ in distribution as $n \to \infty$ for some random variable $V$ with $\mathsf{P}(0 < V < \infty) = 1$. See also a counterexample in Remark 2.

When the FPC scores $\xi_j$ in (2.4) under the model are independent, one can show that $h_n^{-1} t_{h_n}(X_0) \xrightarrow{\mathsf{P}} 1$. In this special case, values of $t_{h_n}(X_0)$ or $h_n$ are equivalent, and the CLT from [CMS] will hold. However, scaling by $t_{h_n}(X_0)$ is generally required for the projection-type CLT in Theorem 1. If the regressor $X$ has the FPC scores $\xi_j$ in (4) that are *dependent*, then Theorem 2 of [CMS] with the scaling factor $h_n$ may not apply. A counterexample is given next, and others are provided in Section 2.10 of the supplement.

**Proposition 1.** *Under the FLRM (2.5), suppose that Conditions (A1)-(A6) hold along with $h_n^{-1} + n^{-1/2} h_n^{5/2} (\log h_n)^2 \to 0$. In addition, suppose that $X$ has FPC scores in (2.4) being $\xi_j = W_j \xi$, $j = 1, 2, \dots$, with an iid sequence $\{W_j\}$ independent of $\xi \sim \mathsf{N}(0,1)$, where $\mathsf{P}(W_1 = 1) = 1/2 = \mathsf{P}(W_1 = -1)$. Then, for a random $X_0$ sharing the same distribution with $X$,*

$$\sqrt{\frac{n}{h_n}} \left[ \langle \hat{\beta}_{h_n}, X_0 \rangle - \langle \Pi_{h_n} \beta, X_0 \rangle \right] \xrightarrow{d} |\xi_0| Z_0, \quad Z_0 \sim \mathsf{N}(0, \sigma_\varepsilon^2),\ \xi_0 \sim \mathsf{N}(0,1),$$

*holds as $n \to \infty$, where $\xi_0$ and $Z_0$ are independent variables.*

The (counter-)example in Proposition 1 serves to show the scaling $t_{h_n}(X_0)$ works in the CLT while the scaling $h_n$ fails. Figure 2.1 provides a numerical illustration, based on 1000 experiments from a FLRM with regressors $X$ generated according to setting described in Proposition 1 (and those described in Section 2.5 with uniform errors and $a = b = 5$). The figure shows that the CLT holds for projections with the updated scaling as in Theorem 1, while the CLT fails with scaling $\sqrt{n/h_n}$, so Theorem 2 of [CMS] does not hold for the setup described in Proposition 1.

*Remark* 2. A reviewer suggested an example based on the model in Proposition 1 with $\xi_j = B Z_j$ where $B$ follows a Bernoulli distribution is independent from $Z_j$ normally distributed. In this case, the condition $h_n t_{h_n}(X_0)^{-1} = O_P(1)$ used in the CLTs of Theorem 1–Corollary 1 fails to hold. Further, we then have $1/2 = P(B = 0) \leq P(t_{h_n}(X_0) = 0)$, which means that the target quantity $\sqrt{n/t_{h_n}}[\langle \hat{\beta}_{h_n}, X_0 \rangle - \langle \beta, X_0 \rangle]$ is not well-normalized with positive probability. This example helps to further motivates the condition $h_n t_{h_n}^{-1}(X_0) = O_{\mathsf{P}}(1)$.

*Remark* 3. [CMS] originally considered a more general way of regularization of $\Gamma$ by using a sequence $\{f_n\}_{n=1}^{\infty}$ of positive functions. If $f_n$ is set to be a reprocical, i.e., $f_n(x) = x^{-1}$, the FPCR

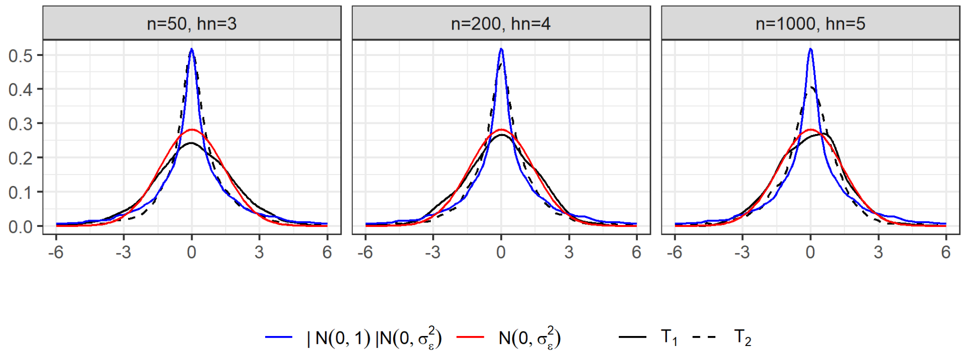


Figure 2.1: Kernel density estimates of $T_1 \equiv \{n/t_{h_n}(X_0)\}^{1/2}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ with scaling $t_{h_n}(X_0)$ (solid black line, according to our result) and $T_2 \equiv \{n/h_n\}^{1/2}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ with scaling $h_n$ (dashed black line, according to [CMS]) over different sample sizes $n$ and different truncation levels $h_n$ depending on sample sizes. The theoretical limits $\mathsf{N}(0, \sigma_\varepsilon^2 = 2)$ of $T_1$ and $|\xi_0|Z_0$ of $T_2$ in Proposition 1 are given for reference (red and blue solid lines, respectively). Centering $\langle\beta, X_0\rangle$ is used for illustration; results for biased centering $\langle\Pi_{h_n}\beta, X_0\rangle$ are similar.

estimator in [CMS] is equal to the FPCR estimator $\hat{\beta}_{h_n}$ in (2.6) as explained in Example 1 therein. The scaling term $s_n$ considered in their Theorem 2 is then exactly the same as the truncation level $h_n$.

Theorem 1 serves to unify the scaling needed for the projection CLT across the cases where the target regressor $X_0$ may be random or conditionally given. In contrast, [CMS] suggests a scaling of $\sqrt{n/h_n}$ for random $X_0$ and $\sqrt{n/t_{h_n}(X_0)}$ for fixed $X_0$. Our results show that scaling $\sqrt{n/h_n}$ is not generally valid in the former case (Proposition 1) and regardless of how $X_0$ is considered, a common scaling $t_{h_n}(X_0)$ or $\hat{t}_{h_n}(X_0)$ should be used. Furthermore, Theorem 1 considerably strengthens the CLT for projections in FLRMs, because this CLT holds conditionally on any given data regressors $\mathcal{X}_n \equiv \{X_i\}_{i=1}^n$, rather than unconditionally as intended in [CMS]. This feature is relevant for the residual bootstrap which can target conditional sampling distributions for $\langle\hat{\beta}_{h_n}, X_0\rangle$ given data regressors $\mathcal{X}_n$.

Theorem 1 cannot be deduced from Theorem 3 of [CMS], where the latter considers a fixed new regressor $x$, even though they look similar. Considering the condition $\sup_{j\in\mathbb{N}} \lambda_j^{-1}\langle X_0, e_j\rangle^2 < \infty$ of their Theorem 3 applied to a random $X_0$, it may hold that

$\mathsf{P}(\sup_{j\in\mathbb{N}} \lambda_j^{-1}\langle X_0, e_j\rangle^2 < \infty) = 0$, for example, when $X_0$ is Gaussian. Furthermore, with fixed a $x$, the bias term related to $\langle(\hat{\Pi}_{h_n} - \Pi_{h_n})\beta, x\rangle$ is hard to remove as explained in Remark 5 of [CMS]. Hence, the fixed $x$ design and the random $X_0$ cases are not directly translatable.

## 2.4 Bootstrap method and results

After outlining the residual bootstrap in Section 2.4.1, Section 2.4.2 establishes the method's consistency for approximating the sampling distribution of regression-based projections

$$\{n/\hat{t}_{h_n}(X_0)\}^{1/2}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle],$$

involving estimated scaling $\hat{t}_{h_n}(X_0)$ as well as centering $\langle\beta, X_0\rangle$ at the true parameter $\beta \in \mathbb{H}$. This justifies the bootstrap for inference about $\langle\beta, X_0\rangle$. Section 2.4.3 then describes the bootstrap for simultaneous confidence regions, while Section 2.4.4 establishes bootstrap prediction intervals for new responses $Y_0$.

### 2.4.1 Residual bootstrap in the FLRM

To describe the residual bootstrap in greatest generality, we consider two integer tuning parameters $k_n, g_n$ for constructing bootstrap data $\{(Y_i^*, X_i)\}_{i=1}^n$ to mimic the original observations $\{(Y_i, X_i)\}_{i=1}^n$ from (2.5). These values define estimators $\hat{\beta}_{k_n}, \hat{\beta}_{g_n}$ from $\{(Y_i, X_i)\}_{i=1}^n$, which are akin to $\hat{\beta}_{h_n}$ in (2.6), but serve exclusively to create $\{(Y_i^*, X_i)\}_{i=1}^n$. With the estimator $\hat{\beta}_{k_n} \equiv \hat{\Gamma}_{k_n}^{\dagger}\hat{\Delta}_n$, we obtain residuals $\hat{\epsilon}_i \equiv Y_i - \langle\hat{\beta}_{k_n}, X_i\rangle$, $i = 1, \ldots, n$ having sample mean $\bar{\hat{\varepsilon}}_n \equiv n^{-1}\sum_{i=1}^n \hat{\epsilon}_i$, and define a sample of bootstrap errors $\varepsilon_1^*, \ldots, \varepsilon_n^*$ as iid uniform draws from $\{\hat{\epsilon}_i - \bar{\hat{\varepsilon}}_n\}_{i=1}^n$. The estimator $\hat{\beta}_{g_n} \equiv \hat{\Gamma}_{g_n}^{\dagger}\hat{\Delta}_n$ then plays the role of the true parameter $\beta$ in the bootstrap world, and the bootstrap sample $\{(Y_i^*, X_i)\}_{i=1}^n$ is defined by

$$Y_i^* = \langle\hat{\beta}_{g_n}, X_i\rangle + \varepsilon_i^*, \quad i = 1, \ldots, n,$$

as an analog of (2.5). Note that both original and bootstrap-recreated data share (or be equally conditional on) the same regressors $\{X_i\}_{i=1}^n$. The bootstrap data $\{(Y_i^*, X_i)\}$ then renders a version $\hat{\beta}_{h_n}^*$ of the original data estimator $\hat{\beta}_{h_n}$ based on a common truncation level $h_n$. Selecting a single $k_n = g_n = h_n$ is possible ([16] use $k_n = h_n$ by default), but it is helpful to separate the effects of tuning parameters in the bootstrap re-construction, as is often considered in resampling problems with kernel estimation (cf. [13]).

Let $X_0$ denote a new regressor under the model, independently of the data $\{(Y_i, X_i)\}_{i=1}^n$. For an observed or given value of $X_0$, we estimate the conditional distribution of

$$T_n(X_0) \equiv \sqrt{\frac{n}{t_{h_n}(X_0)}}\big[\langle \hat{\beta}_{h_n}, X_0\rangle - \langle \beta, X_0\rangle\big]$$

with the bootstrap distribution of

$$T_n^*(X_0^*) \equiv \sqrt{\frac{n}{t_{h_n}(X_0^*)}}\big[\langle \hat{\beta}_{h_n}^*, X_0^*\rangle - \langle \hat{\beta}_{g_n}, X_0^*\rangle\big]$$

with $X_0^* = X_0$ fixed. As a strongest result, both distributions of $T_n(X_0)$ and $T_n^*(X_0^*)$ are viewed as conditional on $X_0$ and on the same data regressors $\mathcal{X}_n \equiv \{X_i\}_{i=1}^n$. For a different case where $X_0$ is unobserved and inference is intended about $\langle \beta, X_0\rangle$ as a random projection, the distribution of $T_n(X_0)$ can still be approximated with a bootstrap counterpart $T_n^*(X_0^*)$, with the change that $X_0^*$ is defined by a random draw from $\mathcal{X}_n$; both distributions remain conditional on $\mathcal{X}_n$, though not $X_0$.

### 2.4.2 Validity of residual bootstrap

To frame the bootstrap results to follow, we first provide a reference result on bootstrap validity for a biased target $\langle \Pi_{h_n}\beta, X_0\rangle$, formed by truncating $\beta$ based on a number $h_n$ of FPCs for defining the estimator $\hat{\beta}_{h_n}$ and with $\Pi_{h_n} \equiv \sum_{i=1}^{h_n}(e_j \otimes e_j)$ from (2.3).

Below let $\tilde{\mathsf{P}}$ denote $\mathsf{P}(\cdot|\mathcal{X}_n, X_0)$ or $\mathsf{P}(\cdot|\mathcal{X}_n)$, conditional on data regressors $\mathcal{X}_n \equiv \{X_i\}_{i=1}^n$ with independent $X_0$ as potentially random or given, and denote the bootstrap probability counterpart as $\tilde{\mathsf{P}}^*(\cdot) \equiv P^*(\cdot|\mathcal{X}_n, X_0^* = X_0)$ or $\tilde{\mathsf{P}}^*(\cdot) \equiv P^*(\cdot|\mathcal{X}_n)$, respectively, where $P^*$ is the bootstrap distribution of the bootstrap data $\{(Y_i^*, X_i)\}_{i=1}^n$. Also, let $\hat{\Pi}_{h_n} \equiv \sum_{i=1}^{h_n}(\hat{e}_j \otimes \hat{e}_j)$, based on estimated eigenfunctions $\hat{e}_j$ (cf. Section 2.2.2), denote the sample analog of $\Pi_{h_n} \equiv \sum_{i=1}^{h_n}(e_j \otimes e_j)$, in order to define a bootstrap version $\hat{\Pi}_{h_n}\hat{\beta}_{g_n}$ of the biased parameter $\Pi_{h_n}\beta$ with $\hat{\beta}_{g_n}$ again playing the bootstrap role of $\beta$.

**Theorem 2.** *Along with assumptions of Theorem 1, suppose that the additional bootstrap truncation $k_n$ satisfies $k_n^{-1} + n^{-1/2}k_n^2 \log k_n + n^{-1}\sum_{j=1}^{k_n} \delta_j^{-2} \to 0$. Then, as $n \to \infty$,*

$$\sup_{y\in\mathbb{R}} \left| \tilde{\mathsf{P}}\left(\sqrt{\frac{n}{t_{h_n}(X_0)}}\left[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\Pi_{h_n}\beta, X_0\rangle\right] \leq y\right) \right.$$
$$\left. -\tilde{\mathsf{P}}^*\left(\sqrt{\frac{n}{t_{h_n}(X_0^*)}}\left[\langle\hat{\beta}^*_{h_n}, X_0^*\rangle - \langle\hat{\Pi}_{h_n}\hat{\beta}_{g_n}, X_0^*\rangle\right] \leq y\right) \right| \xrightarrow{\mathsf{P}} 0,$$

*where the above remains valid if $t_{h_n}$ is replaced by the estimated scaling $\hat{t}_{h_n}$ from (2.8).*

Theorem 2 strengthens the main bootstrap finding of [16] for FLRMs, who considered biased centering, with $X_0$ as given, and $\tilde{\mathsf{P}}$ without conditioning on $\mathcal{X}_n \equiv \{X_i\}_{i=1}^n$. The bootstrap operates under the same basic assumptions as in the biased-case CLT for projections (Theorem 1); no extra conditions are needed for the bootstrap truncation $g_n$, while the conditions for $k_n$ are mild to allow consistent estimation of errors in the bootstrap. For perspective on either the CLT (Theorem 1) or bootstrap (Theorem 2) with *biased* centering, note that there is no strict requirement on the common truncation $h_n$ except that $h_n \to \infty$ as sample size $n \to \infty$ under the condition $h_n^{-1} + n^{-1/2}h_n^{5/2}(\log h_n)^2 \to 0$, and $h_n$ can grow quite slowly relative to $n$; for instance, $h_n = O(n^{1/v_h})$ is acceptable for any $v_h > 5$. That is, while these results impose upper bounds on $h_n$, there are no lower growth rates on $h_n$. However, in order to recover an *unbiased* target $\langle\beta, X_0\rangle$ from $\langle\hat{\beta}_{h_n}, X_0\rangle$, the truncation level $h_n$ will, at least intuitively, need to diverge to infinity sufficiently fast as $n \to \infty$ to adequately capture $\beta$ from the approximate mean $\Pi_{h_n}\beta$ of $\hat{\beta}_{h_n}$. This is considered next.

Theorem 3 establishes bootstrap consistency for sampling distributions defined with an unbiased centering $\langle\beta, X_0\rangle$. Additional smoothness conditions for $\beta$ are needed in the spirit of those used by [CMS] to handle $\Pi_{h_n}\beta - \beta$. Smoothness assumptions are also intricately related to the truncation selections, particularly $h_n$ for the original estimator $\hat{\beta}_{h_n}$ of $\beta$ and $g_n$ for bootstrap re-creation $\hat{\beta}_{g_n}$ of $\beta$.

**Theorem 3.** *Under the assumptions of Theorem 2, suppose that for some constants* $u, v > 0$ *to be specified and a function* $m(j,u) \equiv \max\{j^u, \sum_{i=1}^{j} \delta_i^{-2}\}$ *where* $j = 1, 2, \ldots,$

$$\sup_{j\geq 1} \langle \beta, e_j \rangle^2 j^{v-1} m(j,u) < \infty \tag{2.9}$$

*holds, and that either (a) or (b) holds as follows:*

*(a)* $g_n \leq h_n$ *with* $n = O(h_n^v m(h_n, u))$ *for some* $u > 5$, $v > 0$*;*

*(b)* $g_n > h_n$ *with* $h_n/g_n \to 1$, $n^{-1/2} g_n^{7/2} (\log g_n)^2 \to 0$, $n = O(g_n^v m(g_n, u))$ *for some* $u > 7$, $v > 0$.

*Then, as* $n \to \infty$*, the bootstrap is valid for regression estimators* $\langle \hat{\beta}_{h_n}, X_0 \rangle$ *with unbiased centering* $\langle \beta, X_0 \rangle$*:*

$$\sup_{y\in\mathbb{R}} \left| \tilde{\mathsf{P}} \left( \sqrt{\frac{n}{t_{h_n}(X_0)}} \left[ \langle \hat{\beta}_{h_n}, X_0 \rangle - \langle \beta, X_0 \rangle \right] \leq y \right) - \tilde{\mathsf{P}}^* \left( \sqrt{\frac{n}{t_{h_n}(X_0^*)}} \left[ \langle \hat{\beta}^*_{h_n}, X_0^* \rangle - \langle \hat{\beta}_{g_n}, X_0^* \rangle \right] \leq y \right) \right| \xrightarrow{\mathsf{P}} 0,$$

*where the above remains valid if* $t_{h_n}$ *is replaced by the estimated scaling* $\hat{t}_{h_n}$ *from (2.8).*

For inference about $\langle \beta, X_0 \rangle$ directly, Theorem 3 justifies the residual bootstrap, though the choice of a truncation parameter $h_n$ (or $g_n$ if $g_n > h_n$) is more critical than for the biased target $\langle \Pi_{h_n} \beta, X_0 \rangle$ case of Theorem 2. Under the type-(a) assumption in Theorem 3, the bootstrap truncation $g_n$ for re-creating $\beta$ through $\hat{\beta}_{g_n}$ can be flexibly chosen after the choice of $h_n$ (i.e., less than $h_n$); the truncation $g_n$ can also be larger than $h_n$ through the type-(b) assumption, though they are asymptotically equivalent.

To build a better understanding of the truncation and parameter smoothness conditions in Theorem 3, we may also consider a simpler setting with polynomial decay rates on eigengaps $\delta_j \asymp j^{-a}$ (implying $\lambda_j \asymp j^{-a+1}$) and coordinate projections $|\langle \beta, e_j \rangle| \asymp j^{-b}$ for some constants $a > 2$ and $b > 1$ with $a + 2 < 2b$; here and in the following, we write $r_n \asymp s_n$ if $r_n/s_n$ is bounded away from both zero and infinity for generic sequences $r_n$ and $s_n > 0$. Corollary 2 is a special recasting of Theorem 3.

**Corollary 2.** *Under the above polynomial decay rates, suppose (A1)-(A2) hold along with* $h_n t_{h_n}(X_0)^{-1} = O_{\mathsf{P}}(1)$*,* $k_n \to \infty$*, and* $k_n^{v_k} = O(n)$ *for some* $v_k > \max\{4, (2a+1)\}$*. Suppose further that either (a) or (b) holds as follows:*

*(a)* $g_n \leq h_n$ *with* $n \asymp h_n^{v_h}$ *for some* $\max\{5, (2a+1)\} < v_h < a + 2b - 1$*;*

*(b)* $g_n > h_n$ *with* $h_n/g_n \to 1$ *and* $n \asymp g_n^{v_g}$ *for some* $\max\{7, (2a+1)\} < v_g < a + 2b - 1$*.*

*Then, the conclusions of Theorem 3 remain valid.*

*Remark* 4. Theoretical rates of truncation paramters for FPCR estimators similar to $h_n \asymp n^{1/v_h}$ (or $g_n \asymp n^{1/v_g}$) are quite common and appear in previous works on asymptotic theory in FLRMs such as [5, 19].

Corollary 2 entails that the truncation $h_n$ needs to grow in an appropriate range of rates $n^{1/v_h}$ prescribed by the smoothness of $\beta$ and the eigendecay in the regressor covariance $\Gamma$. The conditions of Corollary 2 also support those used in other estimation studies of the slope function $\beta$ [19] and its projection $\langle \beta, X_0 \rangle$ [5]. The theoretically best rate for $h_n$ in the former work [19] is $h_n \asymp n^{1/(a+2b-1)}$ at the upper bound of the range $v_h \in (\max\{5, 2a+1\}, a+2b-1)$ of Corollary 2, whereas the optimal rate in the latter work [5] is contained in this range. That is, estimation of $\langle \beta, X_0 \rangle$ involves a larger $h_n$ compared to slope $\beta$ estimation, indicating that less smoothing is needed for estimation of $\langle \beta, X_0 \rangle$. This latter point is essentially supported by Corollary 2 in that bootstrap inference about $\langle \beta, X_0 \rangle$ similarly requires a sufficiently large $h_n$ in setting the estimator $\hat{\beta}_{h_n}$. In this sense, while the regression problem with FLRM is parametric, the bootstrap here behaves similarly to resampling in classical nonparametric regression (cf. [17, 18]) where bandwidths are likewise chosen to undersmooth due to bias issues. Numerical studies of the bootstrap in Section 2.5 lead to some recipes for selecting truncations (e.g., $h_n, g_n, k_n$), while data-based truncation choices are considered in the data example of Section 2.6.

*Remark* 5. For simplicity, we have focused on presenting the case where a new (independent) regressor point $X_0$ has the same marginal distribution as that of the original data regressors $\mathcal{X}_n \equiv \{X_i\}_{i=1}^n$. However, the bootstrap results here can be extended to the case when $X_0$ does not have the same distribution. For this, we require additional conditions, similar to those of

Section 2.3 (cf. (A1)-(A6)), but applied to an analogous Karhunen–Loève expansion (2.4) for $X_0$ in place of $X$ from the model (2.1). In particular, suppose that the new regressor $X_0$ has a Karhunen–Loève expansion $X_0 = \sum_{j=1}^{\infty} \mu_j \zeta_j e_j$ with eigenfunctions $\{e_j\}$ being the same as those for the regressor $X$ but possibly different eigenvalues $\{\mu_j\}$ and FPC scores $\{\zeta_j\}$. If $\sup_{j\in\mathbb{N}}(\mu_j/\lambda_j) < \infty$, the CLT Theorem 1 still holds, regardless of the distribution of the FPC scores $\{\zeta_j\}$, upon replacing the scaling $t_{h_n}(X_0)$ there with $r_{h_n}(X_0) \equiv \sum_{j=1}^{h_n} \mu_j^{-1}\langle X_0, e_j\rangle^2 = \sum_{j=1}^{h_n} \zeta_j^2$; the latter is asymptotically equivalent to $h_n$ if the FPC scores $\zeta_j$ are independent. The bootstrap Theorems 2-3 results, conditional on $X_0$ (i.e., $X_0^* = X_0$), also hold replacing $t_{h_n}(X_0)$ with $r_{h_n}(X_0)$; this bootstrap essentially approximates $\langle(\hat{\beta}_{h_n} - \beta), X_0\rangle$ with $\langle(\hat{\beta}^*_{h_n} - \hat{\beta}_{g_n}), X_0\rangle$. When independent replications of $X_0$ are further available in this setting, then the scaling $r_{h_n}(X_0)$ can also be estimated as in (2.8), using these regressor replicates in place of the data regressors $\mathcal{X}_n$.

### 2.4.3 Simultaneous intervals based on bootstrap

A benefit of bootstrap inference in FLRMs is that the method extends readily to simultaneous intervals. Let $\mathcal{X}_0 \equiv \{X_{0,j}\}_{j=1}^m$ denote an iid collection of $m \geq 1$ target regressors, independent of the data $\{(Y_i, X_i)\}_{i=1}^n$, which share the same distribution as a model regressor $X$ (though this may be relaxed as in Remark 5). For inference about the collection $\{\langle\beta, X_{0,j}\rangle\}_{j=1}^m$ of $m$ projections simultaneously, we extend the residual bootstrap to approximate the distribution of a maximal quantity as

$$M_n(\mathcal{X}_0) \equiv \max_{1\leq j\leq m} \sqrt{\frac{n}{\hat{t}_{h_n}(X_{0,j})}} \left|\langle\hat{\beta}_{h_n}, X_{0,j}\rangle - \langle\beta, X_{0,j}\rangle\right|.$$

For observed or given values of $\mathcal{X}_0$, the sampling distribution of $M_n(\mathcal{X}_0)$ may be estimated with the bootstrap distribution of

$$M_n^*(\mathcal{X}_0^*) \equiv \max_{1\leq j\leq m} \sqrt{\frac{n}{\hat{t}_{h_n}(X^*_{0,j})}} \left|\langle\hat{\beta}^*_{h_n}, X^*_{0,j}\rangle - \langle\hat{\beta}_{g_n}, X^*_{0,j}\rangle\right|,$$

with fixed $X^*_{0,j} = X_{0,j}$ for $j = 1, \ldots, m$. The distributions of both $M_n(\mathcal{X}_0)$ and $M_n^*(\mathcal{X}_0^*)$ are interpreted as conditional on $\mathcal{X}_0$ and data regressors $\mathcal{X}_n \equiv \{X_i\}_{i=1}^n$.

When values $\mathcal{X}_0 \equiv \{X_{0,j}\}_{j=1}^m$ are unobserved and inference is intended about $\{\langle \beta, X_{0,j}\rangle\}_{j=1}^m$ as random population locations, we estimate the distribution of $M_n(\mathcal{X}_0)$ (unconditional on $\mathcal{X}_0$, but conditional on $\mathcal{X}_n$) with the bootstrap version $M_n^*(\mathcal{X}_0^*)$ defined by $\mathcal{X}_0^* \equiv \{X_{0,j}^*\}_{j=1}^m$ as an iid sample of size $m$ drawn uniformly from $\mathcal{X}_n$. The following result justifies the bootstrap for simultaneous inference.

Let $\tilde{\mathsf{P}}$ denote $\mathsf{P}(\cdot|\mathcal{X}_n, \mathcal{X}_0)$ or $\mathsf{P}(\cdot|\mathcal{X}_n)$ with $\tilde{\mathsf{P}}^*$ as the bootstrap counterpart $P^*(\cdot|\mathcal{X}_n, \mathcal{X}_0^* = \mathcal{X}_0)$ or $P^*(\cdot|\mathcal{X}_n)$.

**Corollary 3.** *Under the assumptions of Theorem 3, the bootstrap is valid for calibrating simultaneous intervals based on the maximum $M_n(\mathcal{X}_0)$: as $n \to \infty$,*

$$\sup_{y\in\mathbb{R}} \left| \tilde{\mathsf{P}}\left(M_n(\mathcal{X}_0) \leq y\right) - \tilde{\mathsf{P}}^*\left(M_n^*(\mathcal{X}_0^*) \leq y\right) \right| \xrightarrow{\mathsf{P}} 0.$$

Hence, by estimating the $(1-\alpha)$ percentile $q_{1-\alpha}$ of the maximum $M_n(\mathcal{X}_0)$ with the quantile $\hat{q}_{1-\alpha}$ of the bootstrap version $M_n^*(\mathcal{X}_0^*)$, we may define a collection of Scheffé-type intervals

$$\langle \hat{\beta}_{h_n}, X_{0,j}\rangle \pm \hat{q}_{1-\alpha}\sqrt{\hat{t}_{h_n}(X_{0,j})/n}, \quad j = 1, \dots, m,$$

that simultaneously cover $\langle \beta, X_{0,j}\rangle$, $j = 1, \dots, m$, with asymptotically guaranteed coverage $1-\alpha$ (conditionally on any data regressors $\mathcal{X}_n \equiv \{X_i\}_{i=1}^n$).

### 2.4.4 Prediction intervals based on bootstrap

The residual bootstrap in FLRMs can also be used to construct intervals for capturing or predicting the value of a future response $Y_0 \equiv \langle \beta, X_0\rangle + \varepsilon_0$ at some new regressor $X_0$. Note that a prediction interval for $Y_0$ depends heavily on the exact distribution of underlying model errors $\varepsilon_0$, which is not true in the case of a confidence interval for $\langle \beta, X_0\rangle$ that may be based on CLT results instead. In this sense, the bootstrap can be attractive for setting prediction intervals without explicit distributional assumptions about model errors.

At some observed values for regressors $\mathcal{X}_0 \equiv \{X_{0,j}\}_{j=1}^m$, we use the bootstrap to simultaneously predict the collection of $m$ future responses

$$Y_{0,j} \equiv \langle \beta, X_{0,j}\rangle + \varepsilon_{0,j}, \quad j = 1, \dots, m,$$

which are independent of the data $\{(Y_i, X_i)\}_{i=1}^n$ and formed by independent pairs $\{(X_{0,j}, \varepsilon_{0,j})\}_{j=1}^m$ under the model (2.1). Natural data-based predictors of $\{Y_{0,j}\}_{j=1}^m$ are given by the location estimators $\hat{Y}_{0,j} \equiv \langle \hat{\beta}_{h_n}, X_{0,j} \rangle$, $j = 1, \ldots, m$, formed from the estimator $\hat{\beta}_{h_n}$ of $\beta$. The bootstrap goal is then to approximate the distribution of the maximal prediction error

$$E_n(\mathcal{X}_0) \equiv \max_{1 \le j \le m} |Y_{0,j} - \hat{Y}_{0,j}|$$

with the distribution of a bootstrap version

$$E_n^*(\mathcal{X}_0^*) \equiv \max_{1 \le j \le m} |Y_{0,j}^* - \hat{Y}_{0,j}^*|,$$

to calibrate simultaneous prediction intervals for $\{Y_{0,j}\}_{j=1}^m$; see [28] for a similar idea regarding single $m = 1$ predictions from time series.

The construction of $E_n^*(\mathcal{X}_0^*)$ applies the bootstrap prescription from Section 2.4.3. Bootstrap data $\{(Y_i^*, X_i)\}_{i=1}^n$ produces a regression estimator $\hat{\beta}_{h_n}^*$ and bootstrap-analog predictions $\hat{Y}_{0,j}^* \equiv \langle \hat{\beta}_{h_n}^*, X_{0,j}^* \rangle$, where we fix $X_{0,j}^* = X_{0,j}$, $j = 1, \ldots, m$ in defining $\mathcal{X}_0^* \equiv \{X_{0,j}^*\}_{j=1}^m$. Bootstrap versions of the new responses $\{Y_{0,j}\}_{j=1}^m$ are defined in the same fashion as the bootstrap sample $\{(Y_i^*, X_i)\}_{i=1}^n$ itself: $Y_{0,j}^* \equiv \langle \hat{\beta}_{g_n}, X_{0,j}^* \rangle + \varepsilon_{0,j}^*$, $j = 1, \ldots, m$, using $\{\varepsilon_{0,j}^*\}_{j=1}^m$ as iid draws from centered residuals (cf. Section 2.4.2).

Corollary 4 justifies the bootstrap for prediction intervals. Because neither the quantity $E_n(\mathcal{X}_0)$ nor the error terms in (2.1) may have continuous distributions, we state bootstrap convergence in terms of the Levy metric, say $d_L[E_n(\mathcal{X}_0), E_n^*(\mathcal{X}_0^*)|\mathcal{X}_n]$, between the distributions of $E_n(\mathcal{X}_0)$ and $E_n^*(\mathcal{X}_0^*)$, conditional on the data regressors $\mathcal{X}_n$.

**Corollary 4.** *Suppose that the assumptions of either Theorem 3 or Corollary 2 hold. Then, as $n \to \infty$, the bootstrap is consistent for the maximal prediction error: $d_L[E_n(\mathcal{X}_0), E_n^*(\mathcal{X}_0^*)|\mathcal{X}_n] \xrightarrow{\mathsf{P}} 0$.*

Simultaneous prediction intervals via bootstrap are then similar to the simultaneous confidence intervals described in Section 2.4.3, i.e., if $\hat{u}_{1-\alpha}$ denotes the $(1 - \alpha)$ percentile of the bootstrap quantity $E_n^*(\mathcal{X}_0^*)$, then a set of simultaneous prediction intervals for $\{Y_{0,j}\}_{i=1}^m$ is given by $\langle \hat{\beta}_{h_n}, X_{0,j} \rangle \pm \hat{u}_{1-\alpha}$, $j = 1, \ldots, m$.

## 2.5 Simulation studies

Section 2.5.1 describes the design of simulation studies to examine the coverage and width properties of confidence intervals (CIs) and prediction intervals (PIs). For clarity, Section 2.5.2 summarizes findings with pointwise intervals and also provides a rule of thumb for selecting truncations $(h_n, g_n, k_n)$ with the bootstrap. Section 2.5.3 then treats simultaneous intervals.

### 2.5.1 Simulation design

To examine intervals, random samples $\{(Y_i, X_i)\}_{i=1}^n$ of size $n = 50, 200, 1000$ were generated from the FLRM (2.1). Iid regressors $\mathcal{X}_n \equiv \{X_i\}_{i=1}^n$ were generated as random curves on $[0,1]$ along with $n$ iid errors $\{\varepsilon_i\}_{i=1}^n$ having a uniform distribution $U(-a, a)$ for $a = \sqrt{6}$, independently of the regressors. Results for other error distributions (e.g., normal, $t$) were similar; see Section 2.10 of the supplement for more details. Each regressor curve was simulated from a truncated Karhunen–Loève expansion

$$X \stackrel{\mathrm{d}}{=} \sum_{j=1}^{J} \sqrt{\lambda_j}\xi_j e_j \tag{2.10}$$

with $J = 15$; eigengaps having a polynomial decay rate $\delta_j = 3j^{-a}$ of $a$ for $j \geq 1$; and basis functions $\{e_j\}_{j=1}^J$ as the first $J$ functions from the trigonometric basis $\{1, \cos(2\pi x), \sin(2\pi x), \dots\}$. All curves were evaluated at 100 equally spaced points in $[0,1]$. In (2.10), we also used uncorrelated (but dependent) sequences $\{\xi_j\}$ of FPC scores defined as follows: let $\xi_j = V_j W_j$, where $\{W_j\}$ are iid $\mathsf{N}(0,1)$ variables and, independently, let $\{V_j\}$ be a stationary autoregressive process such that each $V_j \sim \mathsf{N}(0,1)$ and $V_{j+1}|V_j \sim \mathsf{N}(0.5V_j, 1.5)$. The slope function $\beta$ was constructed as $\beta = \sum_{j=1}^J w_{\beta,j}|\beta_j|e_j$, with $|\beta_j| = 2j^{-b}$ following a polynomial decay rate of $b$ and with fixed coefficients $w_{\beta,j}$ defined by a initial random draw of $J$ values from $\{-1, 1\}$. Level combinations $(a, b)$ were considered for the different polynomial rates with $a \in \{2.5, 5\}$ and $b \in \{2, 5\}$. Note that all scenarios except $a = 5$ and $b = 2$ satisfy the conditions of Corollary 2 (e.g., $2a + 1 > a + 2b$) for guaranteeing the consistency of the bootstrap. In each scenario, the tuning parameters $h_n$ and $g_n$ were varied in the range $\{1, \dots, 15\}$ to investigate their effects on coverage, while the less consequential bootstrap truncation $k_n$ was chosen as $k_n = 2[n^{1/v_k}]$ with $v_k = 2a + 1 + \kappa_k$ for a small $\kappa_k = 0.1$. For a given data set, bootstrap distributions were approximated from 1000 Monte Carlo resamples.

### 2.5.2 Empirical coverage probabilities


Figure 2.2: Empirical coverages of intervals from bootstrap and normal approximation over different truncations when $n = 50$, $b = 2$. The three columns display the CI for $\langle \Pi_{h_n}\beta, X_0\rangle$, the CI for $\langle \beta, X_0\rangle$, and the PI for $Y_0$, respectively. Crosses $\times$ indicate bootstrap coverages with $h_n, g_n$ selected by a proposed rule.

We next compare CIs and PIs from bootstrap to those from CLT/normal approximations. For each simulation run, an additional regressor $X_0$ was generated independently of the data. Both bootstrap and normal theory CIs were computed for biased targets $\langle \Pi_{h_n}\beta, X_0\rangle$ that vary with truncation $h_n$ defining $\hat{\beta}_{h_n}$, and for the location $\langle \beta, X_0\rangle$ For each simulation combination, the reported coverages of CIs represent the proportion of those intervals covering $\langle \Pi_{h_n}\beta, X_0\rangle$ or $\langle \beta, X_0\rangle$ over 1000 runs. We likewise determined coverage for bootstrap and normal theory PIs for containing a response $Y_0$ generated at $X_0$ in each simulation run. Normal theory PIs implicitly assume that model errors follow a normal distribution, which represents a common practice but is not true for the simulation results presented here. See the supplement for further algorithmic details on the methods and results with other data generations. A nominal level of $1 - \alpha = 0.95$ was used in all scenarios.

We only present results for the smallest sample size $n = 50$ and slowest decay rate $b = 2$ for the slope parameter $\beta$, as the most difficult cases of inference. Additional results and details are included Section 2.10 in the supplement. Figure 2.2 displays the associated coverages of CIs and PIs, where intervals depend on $h_n$ on the horizontal axis; the bootstrap also requires selection of $g_n$, denoted by different lines in Figure 2.2. We observe the following:

1. Bootstrap CIs are always superior to normal-theory CIs in terms of coverage accuracy, provided $h_n$ and $g_n$ are appropriately chosen.

2. Supporting our theory, $g_n$ does not affect bootstrap CIs for biased targets $\langle \Pi_{h_n}\beta, X_0 \rangle$; any reasonably large $h_n$ leads to good coverage.

3. For true projections $\langle \beta, X_0 \rangle$, bootstrap CIs depend on both $g_n$ and $h_n$. While our Theorem 3 allows cases where $g_n$ may be either larger or smaller than $h_n$, the simulation results indicate that only choices with $g_n \leq h_n$ are practically relevant. Setting $h_n = g_n$ worked well for larger $g_n$ but setting $h_n$ to be slightly larger than $g_n$ seems overall preferable for performance, particularly for small $g_n = 1$ or 3.

4. Bootstrap PIs behave similarly to bootstrap CIs and perform much better than normal theory-based PIs; the latter perform especially poorly due to underlying non-normal model errors.

In all, we recommend setting $h_n$ to be slightly larger than $g_n$ in practice. Based on the simulation results in all considered scenarios, we also propose a rule of thumb for selecting tuning parameters $h_n$ and $g_n$ in terms of $k_n$, namely $h_n = [2.21k_n]$ and $g_n = [1.36k_n]$, where $[\cdot]$ denotes the nearest integer. The scaling factors were determined by a linear regression of appropriately

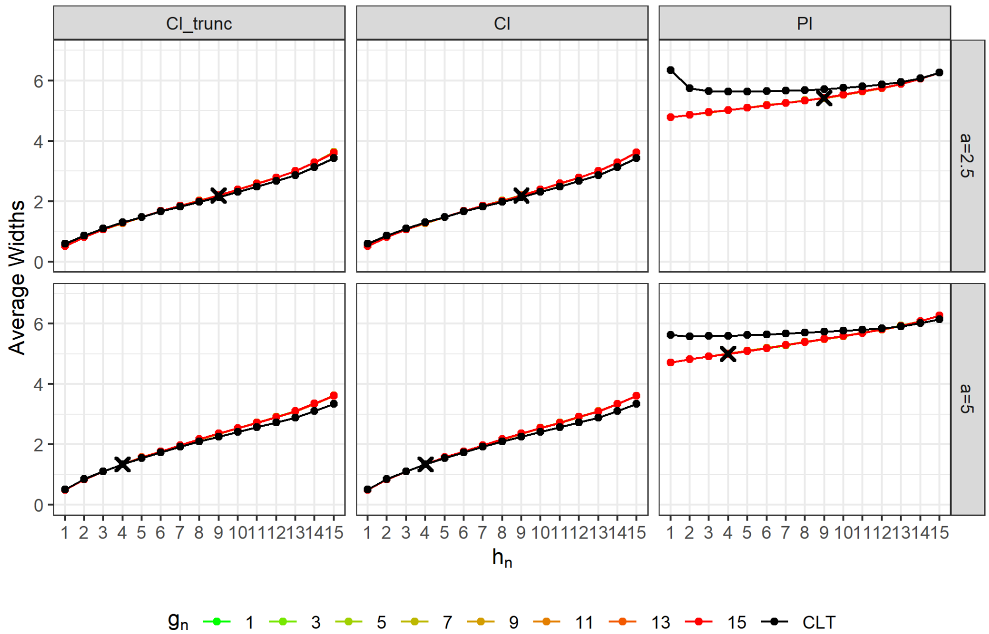


Figure 2.3: Average widths of intervals from bootstrap and normal approximation over different truncations when $n = 50$, $b = 2$. Crosses $\times$ indicate average widths with $h_n, g_n$ selected by the proposed rule. Varying $g_n$ had a negligible effect on the average widths

chosen $k_n$ vs. $(h_n, g_n)$ through manual inspection over all combinations of the latter, from all simulation scenarios, producing coverages within 0.01 of the nominal 95% level. Our rule of thumb exhibited good coverages in Figure 2.2, as indicated by crosses there. In practice, one may appy this procedure with $k_n$ chosen via cross-validation, for example, based on prediction error (cf. Section 2.11 of the supplement).

Figure 2.3 also displays the average widths of intervals, where the rule of thumb tended to produce relatively short- to moderate-width of intervals. Widths of all intervals depend on $h_n$, and generically increase with $h_n$, but bootstrap truncation $g_n$ does not impact widths. However, $g_n$ does impact the coverages of the CIs/PIs, as demonstrated in Figure 2.2, through affecting the "centering" of the bootstrap estimates. Despite having similar average widths in Figure 2.3, bootstrap intervals can outperform normal intervals because the bootstrap better approximates the sampling distribution than the CLT.

### 2.5.3 Coverage of simultaneous intervals

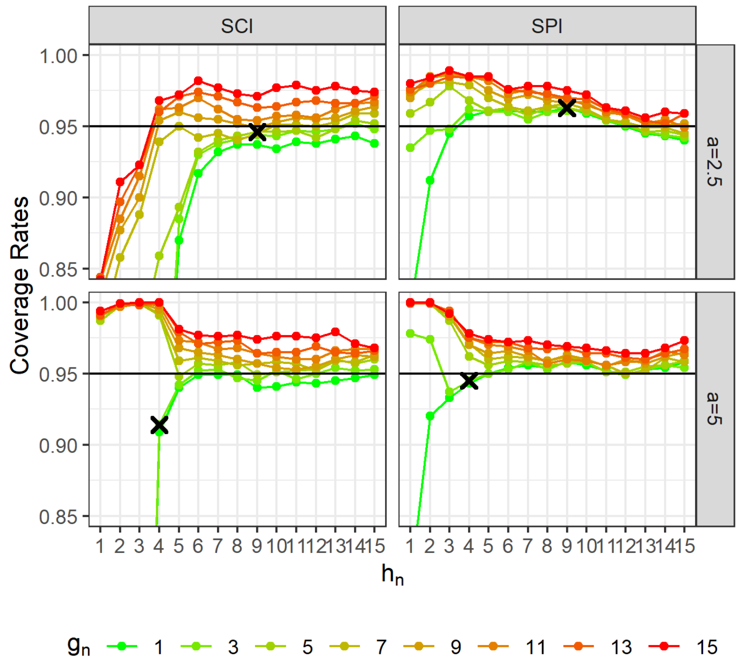


Figure 2.4: Empirical coverages of SPIs and SPIs from bootstrap over different truncations when $n = 50$, $b = 2$, and regressors $\mathcal{X}_0$ are fixed. Crosses $\times$ indicate bootstrap coverages with $h_n, g_n$ selected by the proposed rule.

We next examine simultaneous confidence intervals (SCIs) and prediction intervals (SPIs) via bootstrap, as based on Corollaries 3–4. We focus purely on bootstrap as there is no simple analog for comparison from normal theory. For simplicity, we consider a collection of five regressors $\mathcal{X}_0 = \{X_{0,l}\}_{l=1}^5$ defined by the first five eigenfunctions $X_{0,l} = e_l$, $l = 1, \ldots, 5$, which remained fixed for the study. From the data generated in each simulation run, SCIs were computed for the locations $\{\langle \beta, X_{0,l} \rangle\}_{l=1}^5$, while SPIs were computed for new responses $\{Y_{0,l}\}_{l=1}^5$ at the regressors $\mathcal{X}_0$. Coverage probabilities, as averaged over 1000 simulation runs, were calculated analogously to those in Section 2.5.2 and we likewise present results for the case $n = 50$ with $b = 2$. The supplement summarizes results for other simulation settings, including comparisons to individual CIs/PIs and cases of random regressors. Along lines suggested in Remark 5, note that each regressor $X_{0,l}$ can be viewed as a realization of $\sum_{j=1}^{\infty} \sqrt{\lambda_j}\xi_j e_j$ with FPC scores as $\xi_j = \lambda_l^{-1/2}\mathbb{I}(j = l)$, $j \geq 1$.

Figure 2.4 shows the coverage rates of SCIs/SPIs from bootstrap. Particularly under fixed regressors here, simultaneous intervals tend to exhibit over-coverage, though coverages are often close to nominal for a variety of truncations $h_n, g_n$. However, an important take-away is overly small values of $g_n$ might naturally be avoided, as these can induce extreme under-coverage in SCIs/SPIs due to issues in capturing bias across several intervals at once. Coverage rates from the rule of thumb selections of $h_n$ and $g_n$ given in Section 2.5.2 continue to appear reasonable for SCIs/SPIs in Figure 2.4.

## 2.6 Real Data Analysis

We demonstrate application of the residual bootstrap for FLRMs with a wheat spectrum dataset `Moisturespectrum` from the package `fds`. The dataset, originally described in [22], contains the near-infrared (NIR) reflectance spectra of 100 wheat samples, measured in 2 nm intervals from 1100 to 2500 nm, as well as a response variable, namely the moisture content.

The regressor $X_i$ we analyzed was the negative log-transformed absorption rates $-\log(R(t))$, where $R(t) \in (0, 1)$ denotes the absorption rate at wavelength $t$ over the spectrum $[1100, 2500]$, and the response was the associated moisture level $Y_i$, $i = 1, \ldots, n$. The functional regressors $X_i$

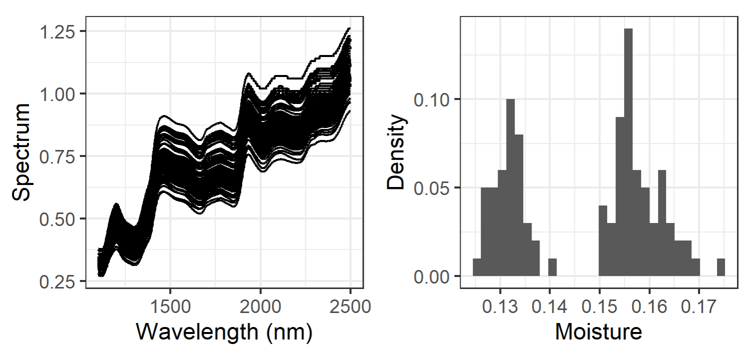


Figure 2.5: The uncentered NIR spectra predictor curves and the distribution of the moisture response

appear in Figure 2.5 along with a distributional summary of the observed responses $Y_i$. After obtaining centered observations $X_i^c = X_i - \bar{X}_n$ and $Y_i^c = Y_i - \bar{Y}_n$, we apply a FLRM (2.1) and estimate the slope parameter $\beta$.

Figure 2.6(a) and (b) show that the distributions of the first two FPC scores do not resemble normal distributions. Additionally, the joint distribution of the first two FPCs appear to follow a slanted v-shape, as shown in Figure 2.6(c), and thus the two FPCs are not independent. For these data, neither Gaussianity nor independence assumption seems reasonable for the FPC scores. However, bootstrap inference is still applicable as per our theoretical results.

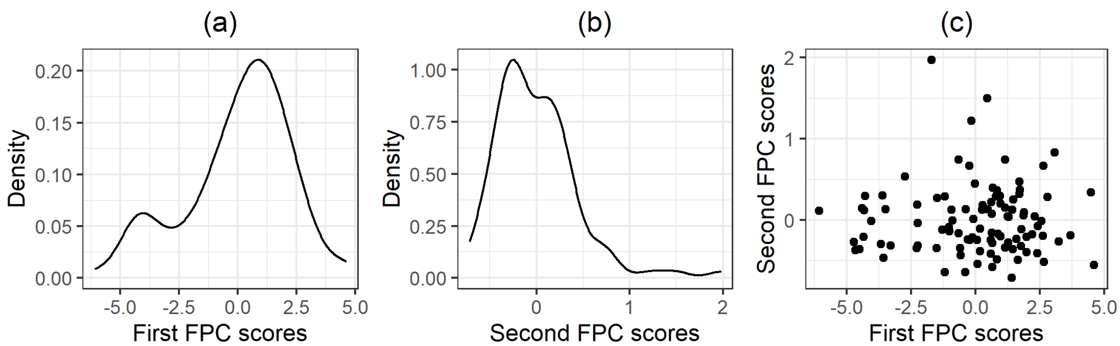


Figure 2.6: Distributions of the first (a) and second (b) FPC scores, and the scatterplot of the first versus the third FPC scores (c)

An initial bootstrap truncation parameter $k_n = 4$ was selected via repeated cross-validation, minimizing prediction errors over estimates from $\hat{\beta}_{k_n}$; the details are included in Section 2.11 of the supplement. Using the selection rule from in Section 2.5.2, we then set $h_n = 9$ and $g_n = 5$.

To illustrate bootstrap-based inference conditional on target regressors $\mathcal{X}_0$, we consider a collection of six hypothetical regressors $\mathcal{X}_0 \equiv \{X_{0,l}\}_{l=1}^6$ of interest and create bootstrap intervals for estimating the true projection, as well as predicting a new response, at these $\mathcal{X}_0$. Three types of regressor collections $\mathcal{X}_0$ are considered: $(OS)$ an overall shift in the magnitude; $(sim)$ a simple functions supported on either [1100 nm, 1400nm], [1400 nm, 1900 nm], or [1900 nm, 2500 nm]; and $(SS)$ a sum of two simple functions in $(sim)$. The $\mathcal{X}_0$ consists of six of a given type, where we vary the type. Types $(sim)$ and $(SS)$ are inspired by two locations where the estimated slope

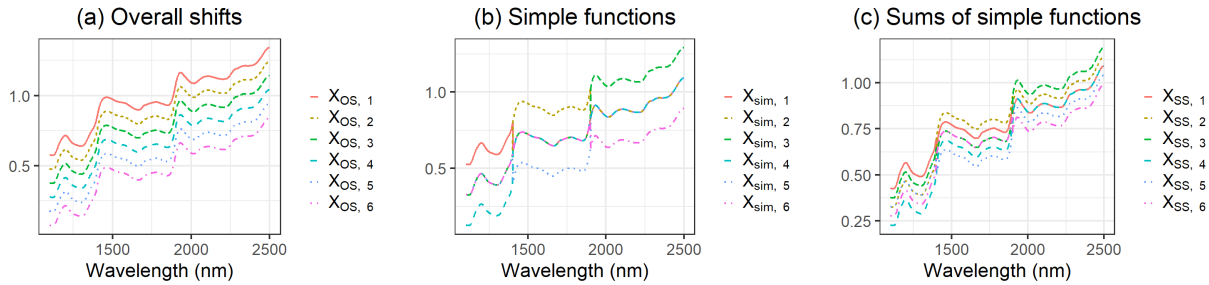


Figure 2.7: Three collections of six hypothetical regressors (curves) for consideration; curves in (b)-(c) partially overlap along the average curve $\bar{X} \equiv n^{-1} \sum_{i=1}^{n} X_i$

function seems to have a peak. Figure 2.7 shows the curves in each set $\mathcal{X}_0$ under consideration after de-centering. The supplement provides more specific details on these curves and the estimated slope. Using bootstrap, Table 2.1 gives the endpoints for 95% individual confidence intervals (ICIs) and simultaneous confidence intervals (SCIs) for $\{\langle \beta, X_{0,l} \rangle\}_{l=1}^{6}$ along with individual prediction intervals (IPIs) and simultaneous prediction intervals (SPIs) for new responses at the regressors $\mathcal{X}_0$. To facilitate comparison to SCIs/SPIs, the ICIs/IPIs are symmetric.

As expected, PIs are wider than the corresponding CIs. From Table 2.1(a), a higher NIR absorption rate, i.e., a curve with lower overall magnitude, is associated with slightly less moisture content. More interesting patterns are involved in the shape of curves or where their peaks are located, which can be seen in Table 2.1(b)-(c). The target regressors which have a peak in the interval [1400 nm, 1900 nm], such as $X_{sim,2}$, $X_{SS,2}$, and $X_{SS,6}$ provide the intervals that contain the lowest value of moisture levels. In contrast, for target regressors having a trough on this interval, for example, $X_{sim,5}$, $X_{SS,3}$, and $X_{SS,5}$, their intervals contains the largest moisture levels. We notice that as target regressors have more pronounced peaks or troughs on this interval, their corresponding intervals reflect lower or higher values of moisture level, respectively.

We note that the ICIs for the first and the sixth regressors among the overall-shift ($OS$) type in Table 2.1(a) match their corresponding SCIs, while the remaining regressors have wider SCIs than PIs. This is not surprising because regressors of the $OS$ type lie in the same one-dimensional space, and the bootstrap procedure automatically accounts for this aspect and calibrates the SCIs according to the two most extreme target regressor curves. It is evident that bootstrap intervals are the widest for the category of simple functions ($sim$), showing that a change in the shapes of regressors has a stronger effect on the response than a change in the overall magnitude.

Table 2.1: ICIs/IPIs and SCIs/SPIs from bootstrap at six target regressors within one of three regressor types ($OS$), ($sim$), and ($SS$).

(a) Overall shifts

| | ICI | | SCI | | IPI | | SPI | |
|---|---|---|---|---|---|---|---|---|
| $X_{OS,1}$ | 17.65 | 19.97 | 17.65 | 19.97 | 17.62 | 20.00 | 17.44 | 20.18 |
| $X_{OS,2}$ | 17.50 | 18.89 | 17.50 | 18.89 | 17.35 | 19.04 | 16.86 | 19.56 |
| $X_{OS,3}$ | 17.34 | 17.81 | 17.34 | 17.81 | 17.04 | 18.11 | 16.21 | 18.94 |
| $X_{OS,4}$ | 16.73 | 17.19 | 16.73 | 17.19 | 16.46 | 17.46 | 15.59 | 18.33 |
| $X_{OS,5}$ | 15.64 | 17.04 | 15.64 | 17.04 | 15.47 | 17.21 | 14.97 | 17.71 |
| $X_{OS,6}$ | 14.56 | 16.88 | 14.56 | 16.88 | 14.38 | 17.07 | 14.35 | 17.09 |

(b) Simple functions

| | ICI | | SCI | | IPI | | SPI | |
|---|---|---|---|---|---|---|---|---|
| $X_{sim,1}$ | 18.34 | 23.47 | 18.23 | 23.57 | 18.43 | 23.37 | 16.58 | 25.21 |
| $X_{sim,2}$ | 8.24 | 16.67 | 7.64 | 17.26 | 8.17 | 16.73 | 8.14 | 16.77 |
| $X_{sim,3}$ | 16.92 | 22.44 | 16.32 | 23.04 | 16.53 | 22.83 | 15.37 | 24.00 |
| $X_{sim,4}$ | 11.06 | 16.21 | 10.96 | 16.31 | 10.92 | 16.35 | 9.32 | 17.95 |
| $X_{sim,5}$ | 17.86 | 26.30 | 17.27 | 26.89 | 17.86 | 26.30 | 17.77 | 26.39 |
| $X_{sim,6}$ | 12.09 | 17.61 | 11.49 | 18.21 | 11.80 | 17.90 | 10.54 | 19.16 |

(c) Sums of two simple functions

| | ICI | | SCI | | IPI | | SPI | |
|---|---|---|---|---|---|---|---|---|
| $X_{SS,1}$ | 17.27 | 18.49 | 17.25 | 18.51 | 17.14 | 18.62 | 15.74 | 20.02 |
| $X_{SS,2}$ | 14.05 | 16.88 | 13.85 | 17.08 | 14.12 | 16.81 | 13.32 | 17.60 |
| $X_{SS,3}$ | 17.37 | 21.39 | 17.14 | 21.63 | 17.40 | 21.36 | 17.24 | 21.52 |
| $X_{SS,4}$ | 16.05 | 17.26 | 16.02 | 17.29 | 16.01 | 17.30 | 14.51 | 18.79 |
| $X_{SS,5}$ | 17.65 | 20.49 | 17.45 | 20.69 | 17.56 | 20.58 | 16.93 | 21.21 |
| $X_{SS,6}$ | 13.14 | 17.16 | 12.90 | 17.40 | 13.04 | 17.26 | 13.01 | 17.29 |

## 2.7 Proof of the generalized/refined CLT

The proof of the CLT is based on the decomposition of $\hat{\beta}_{h_n} - \beta$:

$$\hat{\beta}_{h_n} - \beta = (\hat{\Gamma}^{\dagger}_{h_n} - \Gamma^{\dagger}_{h_n})U_n + \Gamma^{\dagger}_{h_n}U_n + (\hat{\Pi}_{h_n} - \Pi_{h_n})\beta + \Pi_{h_n}\beta - \beta, \tag{2.11}$$

where $U_n \equiv n^{-1}\sum_{i=1}^{n} X_i\varepsilon_i$. The main difference between our results and those of [CMS] is the distributional convergence of the variance term, which is based on $\Gamma^{\dagger}_{h_n}U_n$ in (2.11). Lemma 8 of [CMS] discussed this convergence, but the statement and the proof require some clarification. We establish the following proposition, which refines their Lemma 8.

**Proposition 2.** *Suppose that Condition (A2) and $h_n t_{h_n}(X_0)^{-1} = O_{\mathsf{P}}(1)$ hold. As $n \to \infty$, if $n^{-1}h_n^2 \to 0$, then we have*

$$\sup_{y\in\mathbb{R}} \left| \mathsf{P}\left( \sqrt{\frac{n}{t_{h_n}(X_0)}} \langle \Gamma^{\dagger}_{h_n}U_n, X_0\rangle \le y \middle| \mathcal{X}_n, X_0 \right) - \Phi(y/\sigma_\varepsilon) \right| \xrightarrow{\mathsf{P}} 0.$$

*Proof.* Let $\mathsf{E}^X$ denote conditional expectation given $\mathcal{X}_n, X_0$ in the following. Note that $\sqrt{\frac{n}{\sigma_\varepsilon^2 h_n}}\langle \Gamma^{\dagger}_{h_n}U_n, X_0\rangle = \sum_{i=1}^{n} Z_{i,n}$ where

$$Z_{i,n} = \sigma_\varepsilon^{-1} n^{-1/2} h_n^{-1/2} \langle \Gamma^{\dagger}_{h_n}X_i, X_0\rangle \varepsilon_i.$$

Then, $\mathsf{E}^X[Z_{i,n}] = \sigma_\varepsilon^{-1} n^{-1/2} h_n^{-1/2} \langle \Gamma^{\dagger}_{h_n}X_i, X_0\rangle \mathsf{E}^X[\varepsilon_i] = 0$ holds by independent errors. Set $v_n^2 = \sum_{i=1}^{n} \mathsf{E}^X[Z_{i,n}^2]$. Then, we may write

$$\begin{aligned}
v_n^2 &= n^{-1}h_n^{-1}\sum_{i=1}^{n}\langle X_i, \Gamma^{\dagger}_{h_n}X_0\rangle^2 \\
&= n^{-1}h_n^{-1}\sum_{i=1}^{n}\langle (X_i\otimes X_i)\Gamma^{\dagger}_{h_n}X_0, \Gamma^{\dagger}_{h_n}X_0\rangle \\
&= h_n^{-1}\langle \hat{\Gamma}_n\Gamma^{\dagger}_{h_n}X_0, \Gamma^{\dagger}_{h_n}X_0\rangle \\
&= h_n^{-1}\{A_n + t_{h_n}(X_0)\}
\end{aligned}$$

where $A_n = \langle (\hat{\Gamma}_n - \Gamma)\Gamma^{\dagger}_{h_n} X_0, \Gamma^{\dagger}_{h_n} X_0 \rangle$. Then, by setting $L_{i,n} = \langle \Gamma^{\dagger}_{h_n} X_i, X_0 \rangle$, we formulate a (conditional on $\mathcal{X}_n$ and $X_0$) Lindeberg's condition as

$$\begin{aligned}\mathcal{L}_n &:= v_n^{-2} \sum_{i=1}^{n} \mathsf{E}^X \left[ Z_{i,n}^2 \mathbb{I}(|Z_{i,n}| > \tau v_n) \right] \\ &\leq \frac{1}{\sigma_\varepsilon^2 v_n^2 n h_n} \sum_{i=1}^{n} L_{i,n}^2 \mathsf{E}^X [\varepsilon_i^2 \mathbb{I}(H_n |\varepsilon_i| > \tau)],\end{aligned}$$

where we will show that $\mathcal{L}_n$ converges to 0 by proving

$$H_n \equiv \frac{\max_{1 \leq i \leq n} |L_{i,n}|}{v_n \sqrt{n h_n}} \xrightarrow{\mathsf{P}} 0 \tag{2.12}$$

We assume (2.12) for now and later verify that $H_n \xrightarrow{\mathsf{P}} 0$ as claimed. To verify the Lindeberg condition, we note that, for a given $\eta > 0$, there exist a positive integer $u = u(\eta) > 0$ such that $f(u) < \eta$ (recalling $f$ in Section 2.2.2) and a positive integer $N$ such that $\mathsf{P}(H_n \geq \tau u^{-1}) < \eta$ for $n \geq N$ by $H_n \xrightarrow{\mathsf{P}} 0$. We treat the conditional Lindeberg term $\mathcal{L}_n$ in two cases, depending on the event $H_n < \tau u^{-1}$. When this event holds, we bound $\mathcal{L}_n$ as

$$\begin{aligned}\mathcal{L}_n \mathbb{I}(H_n < \tau u^{-1}) &\leq \left\{ \frac{1}{v_n^2 n h_n} \sum_{i=1}^{n} L_{i,n}^2 \mathsf{E}^X [\varepsilon_i^2 \mathbb{I}(H_n |\varepsilon_i| > \tau)] \right\} \mathbb{I}(H_n < \tau u^{-1}) \\ &\leq \frac{1}{v_n^2 n h_n} \sum_{i=1}^{n} L_{i,n}^2 \mathsf{E}^X [\varepsilon_i^2 \mathbb{I}(H_n |\varepsilon_i| > \tau) \mathbb{I}(H_n < \tau u^{-1})] \\ &\leq \mathsf{E}[\varepsilon_1^2 \mathbb{I}(|\varepsilon_1| > u) | X_1] \leq f(u) < \eta\end{aligned}$$

using $\sum_{i=1}^{n} L_{i,n}^2 = n h_n v_n^2$ and $\mathbb{I}(H_n |\varepsilon_i| > \tau) \mathbb{I}(H_n < \tau u^{-1}) \leq \mathbb{I}(|\varepsilon_i| > u)$; the complement has probability bounded by

$$\mathsf{P}(\mathcal{L}_n \mathbb{I}(H_n \geq \tau u^{-1}) \geq \varepsilon) \leq \mathsf{P}(H_n \geq \tau u^{-1}) < \eta.$$

Consequently, we find $\mathsf{P}(\mathcal{L}_n \geq 2\eta) < \eta$ holds for $n \geq N$, which verifies the Lindeberg condition $\mathcal{L}_n \xrightarrow{\mathsf{P}} 0$ as $n \to \infty$. Furthermore, by Lemma 7 in the supplement, we have $A_n = O_{\mathsf{P}}(n^{-1/2} h_n^2)$, while $A_n t_{h_n}(X_0)^{-1} = O_{\mathsf{P}}(n^{-1/2} h_n)$ by Condition $h_n t_{h_n}(X_0)^{-1} = O_{\mathsf{P}}(1)$. As $n \to \infty$, if $n^{-1} h_n^2 \to 0$, we have

$$\frac{t_{h_n}(X_0)}{A_n + t_{h_n}(X_0)} = \frac{1}{A_n / t_{h_n}(X_0) + 1} = 1 + o_{\mathsf{P}}(1). \tag{2.13}$$

From this and noting that

$$v_n^{-1}\sum_{i=1}^n Z_{i,n} = \sqrt{\frac{t_{h_n}(X_0)}{A_n + t_{h_n}(X_0)}}\sqrt{\frac{n}{\sigma_\varepsilon^2 t_{h_n}(X_0)}}\langle \Gamma_{h_n}^\dagger U_n, X_0\rangle,$$

we have

$$\sup_{y\in\mathbb{R}}\left|\mathsf{P}\left(v_n^{-1}\sum_{i=1}^n Z_{i,n} \le y \,\middle|\, \mathcal{X}_n, X_0\right) - \Phi(y)\right| \xrightarrow{\mathsf{P}} 0$$

by the Lindeberg CLT and by Polya's theorem (Theorem 9.1.4 of [2]), which is equivalent to

$$\sup_{y\in\mathbb{R}}\left|\mathsf{P}\left(\sqrt{\frac{n}{t_{h_n}(X_0)}}\langle \Gamma_{h_n}^\dagger U_n, X_0\rangle \le y \,\middle|\, \mathcal{X}_n, X_0\right) - \Phi(y/\sigma_\varepsilon)\right| \xrightarrow{\mathsf{P}} 0.$$

We next show that $H_n \xrightarrow{\mathsf{P}} 0$ in (2.12). Due to (2.13), to establish (2.12), it suffices to show

$$M_n \equiv \max_{1\le i\le n} |L_{i,n}| = O_{\mathsf{P}}((nh_n^2)^{1/4})\sqrt{t_{h_n}(X_0)}, \tag{2.14}$$

which then implies

$$\begin{aligned} H_n &= O_{\mathsf{P}}((h_n^2/n)^{1/4})\sqrt{\frac{t_{h_n}(X_0)}{A_n + t_{h_n}(X_0)}} \\ &= O_{\mathsf{P}}((h_n^2/n)^{1/4}), \end{aligned}$$

and $H_n \xrightarrow{\mathsf{P}} 0$ by $n^{-1}h_n^2 \to 0$ as $n \to \infty$. To establish (2.14), note that

$$\begin{aligned} |L_{i,n}| &= \left|\sum_{j=1}^{h_n}\lambda_j^{-1}\langle X_i, e_j\rangle\langle X_0, e_j\rangle\right| \le \sqrt{\sum_{j=1}^{h_n}\lambda_j^{-1}\langle X_i, e_j\rangle^2}\sqrt{\sum_{j=1}^{h_n}\lambda_j^{-1}\langle X_0, e_j\rangle^2} \\ &= \sqrt{\sum_{j=1}^{h_n}\lambda_j^{-1}\langle X_i, e_j\rangle^2}\sqrt{t_{h_n}(X_0)}. \end{aligned}$$

Also, we have by Jensen inequality that

$$\left(\sum_{j=1}^{h_n}\lambda_j^{-1}\langle X_i, e_j\rangle^2\right)^2 \le h_n\sum_{j=1}^{h_n}\left(\lambda_j^{-1}\langle X_i, e_j\rangle^2\right)^2 = h_n\sum_{j=1}^{h_n}\lambda_j^{-2}\langle X_i, e_j\rangle^4.$$

By the finite fourth moment assumption (A2), we see that

$$\mathsf{E}\left[\max_{1\le i\le n}\sum_{j=1}^{h_n}\lambda_j^{-2}\langle X_i, e_j\rangle^4\right] \le \mathsf{E}\left[\sum_{i=1}^n\sum_{j=1}^{h_n}\lambda_j^{-2}\langle X_i, e_j\rangle^4\right] \le nh_nC,$$

which implies that, $\max_{1\le i\le n}\sum_{j=1}^{h_n}\lambda_j^{-2}\langle X_i,e_j\rangle^4 = O_{\mathsf{P}}(nh_n)$. Therefore, we have that

$$M_n \le \left(h_n \max_{1\le i\le n}\sum_{j=1}^{h_n}\lambda_j^{-2}\langle X_i,e_j\rangle^4\right)^{1/4}\sqrt{t_{h_n}(X_0)} = O_{\mathsf{P}}((nh_n^2)^{1/4})\sqrt{t_{h_n}(X_0)},$$

proving (2.14). □

*Proof of Theorem 1.* The theorem follows by applying Proposition 2, Propositions 17-18 in the supplement, the decomposition (2.11), and Polya's theorem (Theorem 9.1.4 of [2]). □

*Proof of Proposition 1.* It suffices to show the weak covergence of the variance term $\sqrt{n/h_n}\langle\Gamma_{h_n}^{\dagger}U_n,X_0\rangle$. For any given regressor $X$, the FPCs $\xi_j$ are uncorrelated with mean zero, variance one, and finite fourth moments, and hence satisfy the assumptions involved. However, these are not independent due to their common component $\xi$. To derive the weak convergence result, we first notice that $h_n^{-1}t_{h_n}(X_0)=\xi_0^2$, where $\xi_0$ denotes the copy of $\xi$ for defining $X_0$, which is independent of the data $\{(Y_i,X_i)\}_{i=1}^n$. Then, by applying Theorem 1 with the bounded convergence theorem, noting also $\mathsf{P}(\xi_0=0)=0$, we find

$$\begin{aligned}\lim_{n\to\infty}\mathsf{P}\left(\sqrt{\frac{n}{h_n}}\langle\Gamma_{h_n}^{\dagger}U_n,X_0\rangle\le y\right) &= \lim_{n\to\infty}\mathsf{E}\left[\mathsf{P}\left(|\xi_0|\sqrt{\frac{n}{t_{h_n}(X_0)}}\langle\Gamma_{h_n}^{\dagger}U_n,X_0\rangle\le y\Big|X_0,\mathcal{X}_n\right)\right]\\ &=\mathsf{E}\left[\Phi\left(\frac{y}{\sigma_\varepsilon|\xi_0|}\right)\right]=\mathsf{P}(|\xi_0|Z_0\le y/\sigma_\varepsilon),\end{aligned}$$

where $Z_0\sim N(0,1)$ denote a random variable independent of $\xi_0$. □

### 2.8 Validity of the residual bootstrap

The Mallow's metric [3] is applied to show that some key distributional components have the same limits in both bootstrap and original data worlds. The Mallows metric, denoted by $d_2$, is a metric between either two distributions on a separable Banach spaces $\mathbb{B}$ or two random variables that can be valued in $\mathbb{B}$. The Mallows metric between two probability distributions $P$ and $Q$ on $\mathbb{B}$ is defined as

$$d_2(P,Q)=\inf_{U'\sim P,V'\sim Q}(\mathsf{E}[\|U'-V'\|^2])^{1/2},$$

where $(U', V')$ denote any pair of two $\mathbb{B}$-valued random variables with $P$ and $Q$ being the marginal distributions of $U'$ and $V'$, respectively. With abuse of notation, the Mallows metric between (the laws of) two random variables $U$ and $V$ taking values in $\mathbb{B}$ is similarly defined by

$$d_2(U, V) = \inf_{U' \overset{\mathrm{d}}{=} U, V' \overset{\mathrm{d}}{=} V} (\mathsf{E}[\|U' - V'\|^2])^{1/2}.$$

See Section 8 of [3] for more details. Let $d_2^X$ denote the Mallows metric defined via the conditional expectation $\mathsf{E}^X$ given $\mathcal{X}_n$ and $X_0$.

We divide the proof of validity of the residual bootstrap into two cases, namely, with or without the bias. Unless otherwise stated, we impose the assumptions in Theorem 2. We first notice the following decomposition of bootstrap quantity

$$\begin{aligned}\hat{\beta}^*_{h_n} - \hat{\beta}_{g_n} &= \hat{\Gamma}^\dagger_{h_n} \hat{\Delta}^*_n - \hat{\beta}_{g_n} = \hat{\Gamma}^\dagger_{h_n} \hat{\Gamma}_n \hat{\beta}_{g_n} + \hat{\Gamma}^\dagger_{h_n} U^*_n - \hat{\beta}_{g_n} \\ &= \hat{\Gamma}^\dagger_{h_n} U^*_n + \hat{\Pi}_{h_n} \hat{\beta}_{g_n} - \hat{\beta}_{g_n}.\end{aligned}$$

Here, $\hat{\beta}^*_{h_n} - \hat{\Pi}_{h_n} \hat{\beta}_{g_n} = \hat{\Gamma}^\dagger_{h_n} U^*_n$ represents the variance part whereas $\hat{\Pi}_{h_n} \hat{\beta}_{g_n} - \hat{\beta}_{g_n}$ is the bias part in the bootstrap world. We hence compare these variance and bias parts between the real and the bootstrap worlds.

Thus, we will show that $\sqrt{\frac{n}{t_{h_n}(X_0)}} \langle \hat{\Gamma}^\dagger_{h_n} U_n, X_0 \rangle$ and $\sqrt{\frac{n}{t_{h_n}(X_0)}} \langle \hat{\Gamma}^\dagger_{h_n} U^*_n, X_0 \rangle$ conditional on $X_1, \ldots, X_n$ and $X_0$ has the same distributional limit by proving the Mallows metric between them converges to zero in probability.

**Proposition 3.** *The Mallows metric between the variance terms conditional on $\mathcal{X}_n$ and $X_0$ satisfies*

$$\begin{aligned}&d_2^X \left( \sqrt{\frac{n}{t_{h_n}(X_0)}} \langle \hat{\Gamma}^\dagger_{h_n} U_n, X_0 \rangle, \sqrt{\frac{n}{t_{h_n}(X_0)}} \langle \hat{\Gamma}^\dagger_{h_n} U^*_n, X_0 \rangle \right)^2 \\ &= \left\{ O_{\mathsf{P}} \left( n^{-1/2} h_n^{-1} \sum_{j=1}^{h_n} (j \log j)^2 \right) + 1 \right\} d_2(F, \hat{F}_n)^2,\end{aligned}$$

*where $F$ and $\hat{F}_n$ denote the distribution functions of errors $\{\varepsilon_i\}_{i=1}^n$ and the centered residuals $\{\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}\}_{i=1}^n$, respectively. Thus, as $n \to \infty$, if $n^{-1/2} h_n^2 (\log h_n)^2 = O(1)$, and $d_2(F, \hat{F}_n) \to 0$, then the conditional Mallows metric converges to zero in probability.*

*Proof.* The proof is along the lines of the proof of Theorem 4.1 in [26]. Since the infimum in the Mallows metric is attained due to Lemma 8.1 of [3], there exists iid $(\varepsilon_i', \varepsilon_i^{*\prime})$'s such that

1. $\varepsilon_i' \sim F$, $\varepsilon_i^{*\prime} \sim \hat{F}_n$,
2. $(\varepsilon_i', \varepsilon_i^{*\prime})$ is independent of $X_i$, and
3. $\mathsf{E}^X[(\varepsilon_i' - \varepsilon_i^{*\prime})^2] = \mathsf{E}[(\varepsilon_i' - \varepsilon_i^{*\prime})^2] = d_2(F, \hat{F}_n)^2$.

Then, we have

$$\begin{aligned} &d_2^X\left(\sqrt{\frac{n}{t_{h_n}(X_0)}}\langle \hat{\Gamma}_{h_n}^{\dagger} U_n, X_0\rangle, \sqrt{\frac{n}{t_{h_n}(X_0)}}\langle \hat{\Gamma}_{h_n}^{\dagger} U_n^*, X_0\rangle\right)^2 \\ \leq &\frac{n}{t_{h_n}(X_0)}\mathsf{E}^X\left[\langle \hat{\Gamma}_{h_n}^{\dagger}(U_n' - U_n^{*\prime}), X_0\rangle^2\right] \end{aligned}$$

where $U_n' = n^{-1}\sum_{i=1}^n X_i\varepsilon_i'$ and $U_n^{*\prime} = n^{-1}\sum_{i=1}^n X_i\varepsilon_i^{*\prime}$. Note that

$$\begin{aligned} \mathsf{E}^X\left[\langle \hat{\Gamma}_{h_n}^{\dagger}(U_n' - U_n^{*\prime}), X_0\rangle^2\right] =&\mathsf{E}^X\left[\left\{n^{-1}\sum_{i=1}^n \langle \hat{\Gamma}_{h_n}^{\dagger} X_i, X_0\rangle(\varepsilon_i' - \varepsilon_i^{*\prime})\right\}^2\right] \\ =&n^{-2}\sum_{i=1}^n \langle \hat{\Gamma}_{h_n}^{\dagger} X_i, X_0\rangle^2 d_2(F, \hat{F}_n)^2. \end{aligned}$$

Also, note that

$$\begin{aligned} n^{-1}\sum_{i=1}^n \langle \hat{\Gamma}_{h_n}^{\dagger} X_i, X_0\rangle^2 &= n^{-1}\sum_{i=1}^n \langle (X_i \otimes X_i)\hat{\Gamma}_{h_n}^{\dagger} X_0, \hat{\Gamma}_{h_n}^{\dagger} X_0\rangle = \langle \Gamma_n \hat{\Gamma}_{h_n}^{\dagger} X_0, \hat{\Gamma}_{h_n}^{\dagger} X_0\rangle \\ &= \langle \hat{\Gamma}_{h_n}^{\dagger} X_0, X_0\rangle = \hat{t}_{h_n}(X_0), \end{aligned}$$

and thus

$$d_2^X\left(\sqrt{\frac{n}{t_{h_n}(X_0)}}\langle \hat{\Gamma}_{h_n}^{\dagger} U_n, X_0\rangle, \sqrt{\frac{n}{t_{h_n}(X_0)}}\langle \hat{\Gamma}_{h_n}^{\dagger} U_n^*, X_0\rangle\right)^2 \leq \frac{\hat{t}_{h_n}(X_0)}{t_{h_n}(X_0)} d_2(F, \hat{F}_n)^2.$$

By Proposition 8 in the supplement, we have $\hat{t}_{h_n}(X_0)/t_{h_n}(X_0) = 1 + O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}(j\log j)^2\right)$, and thus, we conclude that

$$\begin{aligned} &d_2^X\left(\sqrt{\frac{n}{t_{h_n}(X_0)}}\langle \hat{\Gamma}_{h_n}^{\dagger} U_n, X_0\rangle, \sqrt{\frac{n}{t_{h_n}(X_0)}}\langle \hat{\Gamma}_{h_n}^{\dagger} U_n^*, X_0\rangle\right)^2 \\ =&\left\{O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}(j\log j)^2\right) + 1\right\} d_2(F, \hat{F}_n)^2. \end{aligned}$$

□

*Proof of Theorem 2.* We consider here the bootstrap approximation conditional on $X_0^* = X_0$, and the supplement treats the unconditional (on $X_0$) bootstrap case. Write $\hat{v}_n \equiv \sqrt{n/t_{h_n}(X_0)}\langle \hat{\Gamma}_{h_n}^{\dagger} U_n, X_0\rangle$ and

$$\hat{v}_n^* \equiv \sqrt{n/t_{h_n}(X_0)}\langle \hat{\Gamma}_{h_n}^{\dagger} U_n^*, X_0\rangle = \sqrt{n/t_{h_n}(X_0)}[\langle \hat{\beta}_{h_n}^*, X_0\rangle - \langle \hat{\Pi}_{h_n}\hat{\beta}_{g_n}, X_0\rangle].$$

By consistency of the bootstrap error (cf. Theorem 6 in the supplement), under the assumptions in Theorem 2, we have $d_2^X(\hat{v}_n, \hat{v}_n^*) \xrightarrow{\mathsf{P}} 0$ as $n \to \infty$. Meanwhile, to show the convergence of $\hat{v}_n$ in the Mallows metric, note that

$$\begin{aligned}
\mathsf{E}^X[\langle \hat{\Gamma}_{h_n}^{\dagger} U_n, X_0\rangle^2] &= n^{-2}\sum_{i=1}^n \mathsf{E}^X[\langle X_i\varepsilon_i, \hat{\Gamma}_{h_n}^{\dagger} X_0\rangle^2] + n^{-2}\sum_{i\neq i'} \mathsf{E}^X[\langle X_i\varepsilon_i, \hat{\Gamma}_{h_n}^{\dagger} X_0\rangle\langle X_{i'}\varepsilon_{i'}, \Gamma_{h_n}^{\dagger} X_0\rangle] \\
&= \frac{\sigma_\varepsilon^2}{n^2}\sum_{i=1}^n \langle X_i, \hat{\Gamma}_{h_n}^{\dagger} X_0\rangle^2 = \frac{\sigma_\varepsilon^2}{n^2}\sum_{i=1}^n \langle (X_i \otimes X_i)\hat{\Gamma}_{h_n}^{\dagger} X_0, \hat{\Gamma}_{h_n}^{\dagger} X_0\rangle \\
&= \frac{\sigma_\varepsilon^2}{n}\langle X_0, \hat{\Gamma}_{h_n}^{\dagger} X_0\rangle = \frac{\sigma_\varepsilon^2}{n}\hat{t}_{h_n}(X_0),
\end{aligned}$$

which implies that as $n \to \infty$,

$$\mathsf{E}^X[\hat{v}_n^2] = \sigma_\varepsilon^2 \frac{\hat{t}_{h_n}(X_0)}{t_{h_n}(X_0)} \xrightarrow{\mathsf{P}} \sigma_\varepsilon^2$$

by Proposition 8 in the supplement. Therefore, by Lemma 8.3 of [3], Proposition 2, Proposition 17 in the supplement, and Slutksy's theorem (Theorem 9.1.6 of [2]), as $n \to \infty$, we have $d_2^X(\hat{v}_n, Z) \xrightarrow{\mathsf{P}} 0$, which implies that $d_2^X(\hat{v}_n^*, Z) \xrightarrow{\mathsf{P}} 0$, where $Z \sim \mathsf{N}(0, \sigma_\varepsilon^2)$. Therefore, we have the desired result by Theorem 1, Lemma 8.3 of [3], and Polya's theorem (Theorem 9.1.4 of [2]). □

To deal with the bias terms in the real and bootstrap worlds, let $b_n \equiv \sqrt{n/t_{h_n}(X_0)}[\langle \Pi_{h_n}\beta, X_0\rangle - \langle \beta, X_0\rangle]$ and $\hat{b}_n \equiv \sqrt{n/t_{h_n}(X_0)}[\langle \hat{\Pi}_{h_n}\hat{\beta}_{g_n}, X_0\rangle - \langle \hat{\beta}_{g_n}, X_0\rangle]$. The difference between bias terms is

$$\hat{\Pi}_{h_n}\hat{\beta}_{g_n} - \hat{\Pi}_{g_n} = (\hat{\Pi}_{h_n} - \Pi_{h_n})(\hat{\beta}_{g_n} - \beta) + (\hat{\Pi}_{h_n} - \Pi_{h_n})\beta + (\Pi_{h_n} - I)(\hat{\beta}_{g_n} - \beta) + (\Pi_{h_n} - I)\beta. \tag{2.15}$$

*Proof of Theorem 3.* We suppose for now and later verify $\hat{b}_n \xrightarrow{\mathsf{P}} 0$ as $n \to \infty$. Conditional on $X_0$ and following the proof of Theorem 2, we have $d_2^X(\hat{v}_n^*, Z) \xrightarrow{\mathsf{P}} 0$. By using a subsequence argument (cf. Theorem 20.5 of [4]) and Slutksy's theorem (Theorem 9.1.6 of [2]), one can show that

$$\sqrt{\frac{n}{t_{h_n}(X_0)}}[\langle\hat{\beta}^*_{h_n}, X_0\rangle - \langle\hat{\beta}_{g_n}, X_0\rangle] = \hat{v}^*_n + \hat{b}_n \xrightarrow[n\to\infty]{\mathrm{d}} \mathsf{N}(0, \sigma^2_\varepsilon)$$

along the subsequence, pointwise on an almost sure set. In other words, as $n \to \infty$, we have

$$\sup_{y\in\mathbb{R}}\left|\mathsf{P}^*\left(\sqrt{\frac{n}{t_{h_n}(X_0)}}[\langle\hat{\beta}^*_{h_n}, X^*_0\rangle - \langle\hat{\beta}_{g_n}, X^*_0\rangle] \le y|\mathcal{X}_n, X^*_0 = X_0\right) - \Phi(y/\sigma_\varepsilon)\right| \xrightarrow{\mathsf{P}} 0$$

by Polya's theorem (Theorem 9.1.4 of [2]). Using the above, the proof for the unconditional (on $X_0$) bootstrap version in Theorem 3 follows by the argument used for Theorem 2 in this same case.

It suffices to show that $\hat{b}_n \xrightarrow{\mathsf{P}} 0$ as $n \to \infty$. We first suppose that $g_n \le h_n$, which implies that $\hat{b}_n = 0$. Due to Theorem 2, it is enough to show that

$$\sqrt{\frac{n}{t_{h_n}(X_0)}}\langle\Pi_{h_n}\beta - \beta, X_0\rangle \xrightarrow{\mathsf{P}} 0$$

as $n \to \infty$. Thus, the desired result follows from Lemma 9 in the supplement after applying Condition (2.9).

We now suppose that $g_n > h_n$ for the second part. The proof is based on the decomposition (2.15) of the difference between bias terms. Due to Propositions 11-12 in the supplement, if $g_n/h_n \to 1$, one can show that

$$\begin{aligned}&\mathsf{E}^X\left[\sqrt{\frac{n}{t_{h_n}(X_0)}}|\langle(\hat{\Pi}_{h_n} - \Pi_{h_n})(\hat{\beta}_{g_n} - \beta), X_0\rangle|\right]\\ =&O_{\mathsf{P}}\left(M_{n,g_n}h_n^{-1/2}\sum_{j=1}^{h_n} j\log j\right) + O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sqrt{\sum_{j>g_n}\beta_j^2}\sum_{j=1}^{h_n}(j\log j)^2\right),\end{aligned}$$

where for integer $j \ge 1$, $M_{n,j}$ is defined as

$$M_{n,j} = n^{-1}\sum_{l=1}^{j}\delta_l^{-1/2}(l\log l)^{3/2} + n^{-1/2}\left(\sum_{l=1}^{j}\gamma_l^{-1}\right)^{1/2} + n^{-1/2}\sum_{l=1}^{j} l\log l, \tag{2.16}$$

and

$$\begin{aligned}&\mathsf{E}^X\left[\sqrt{\frac{n}{t_{h_n}(X_0)}}|\langle(I - \Pi_{h_n})(\hat{\beta}_{g_n} - \beta), X_0\rangle|\right]\\ =&O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{g_n}(j\log j)^2\right) + O_{\mathsf{P}}\left(\sqrt{\frac{n}{h_n}\sum_{j>g_n}\gamma_j\beta_j^2}\right).\end{aligned}$$

Finally, the result follows again from Lemma 9 in the supplement after applying Condition (2.9), along with the condition $n^{-1/2}g_n^{7/2}(\log g_n)^2 \to 0$. $\square$

### 2.9 Technical details

In this section, we complete the proofs for the main results of the paper. After some preliminary results related to the perturbation theory in functional analysis, we provide proofs for our generalized CLT and the residual bootstrap in Sections 2.3-2.4, respectively, of the main paper.

Recall that $\hat{\beta}_{h_n} - \beta$ is decomposed as follows:

$$\hat{\beta}_{h_n} - \beta = (\hat{\Gamma}^{\dagger}_{h_n} - \Gamma^{\dagger}_{h_n})U_n + \Gamma^{\dagger}_{h_n}U_n + (\hat{\Pi}_{h_n} - \Pi_{h_n})\beta + \Pi_{h_n}\beta - \beta. \tag{2.17}$$

To deal with the bias terms related to $(\hat{\Gamma}^{\dagger}_{h_n} - \Gamma^{\dagger}_{h_n})U_n$ and $(\hat{\Pi}_{h_n} - \Pi_{h_n})\beta$, we apply perturbation theory or functional calculus as seen in many existing works such as Cardot, Mas, and Sarda [7] (referred to as [CMS] hereafter) or [9]. We refer to Chapter VII of [11] or Chapter 5 of [21] for an overview.

Write $\|\cdot\|_\infty$ for the operator norm, namely $\|A\|_\infty \equiv \sup_{\|v\|=1}\|Av\|$. Let $\mathcal{B}_j = \{z \in \mathbb{C} : |z - \lambda_j| = \delta_j/2\}$ be the oriented circle in the complex plane $\mathbb{C}$ and set $\mathcal{C}_n = \bigcup_{j=1}^{h_n}\mathcal{B}_j$ to define the contour integral for operator-valued functions. By the theory from functional calculus (for the bounded linear operators) or perturbation theory, we see that

$$\Pi_{h_n} = \sum_{j=1}^{h_n}\pi_j = \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}(zI - \Gamma)^{-1}dz = \frac{1}{2\pi\iota}\int_{\mathcal{C}_n}(zI - \Gamma)^{-1}dz,$$

$$\Gamma^{\dagger}_{h_n} = \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}z^{-1}(zI - \Gamma)^{-1}dz = \frac{1}{2\pi\iota}\int_{\mathcal{C}_n}z^{-1}(zI - \Gamma)^{-1}dz$$

where $\pi_j = \frac{1}{2\pi\iota}\int(zI - \Gamma)^{-1}dz$ denotes the Riesz projection of $\Gamma$ to corresponding to the $j$-th eigenvalue $\lambda_j$, which is the projection operator onto the $j$-th eigenfunction $e_j$. One can also get the empirical counterparts $\hat{\pi}_j$, $\hat{\Pi}_{h_n}$, and $\hat{\Gamma}^{\dagger}_{h_n}$ from the sample covariance operator $\hat{\Gamma}_n$ with the corresponding random contours $\hat{\mathcal{B}}_j = \{z \in \mathbb{C} : |z - \hat{\lambda}_j| = \hat{\delta}_j/2\}$ and $\hat{\mathcal{C}}_n = \bigcup_{j=1}^{h_n}\hat{\mathcal{B}}_j$. For later development, write random operator-valued functions $G_n(z) = (zI - \Gamma)^{-1/2}(\hat{\Gamma}_n - \Gamma)(zI - \Gamma)^{-1/2}$,

$K_n(z) = (zI - \Gamma)^{1/2}(zI - \hat{\Gamma}_n)^{-1}(zI - \Gamma)^{1/2}$, and event $\mathcal{E}_j = (\|G_n(z)\|_\infty < 1/2, \forall z \in \mathcal{B}_j)$. Due to the frequent uses, we state the following lemma without proof, which corresponds to Lemmas 1-4 in Section 6.1 of [CMS].

**Lemma 1.** *Suppose that Conditions (A2) and (A4) in the main paper hold. We have the following:*

1. *For sufficiently large $j, k \in \mathbb{N}$ with $k < j$, we have*

$$j\lambda_j \geq k\lambda_k, \quad \frac{\lambda_j}{\lambda_j - \lambda_k} \leq \frac{k}{k-j}, \quad \text{and} \quad \sum_{j \geq k} \lambda_j \leq (k+1)\lambda_k.$$

2. *For sufficiently large $j \in \mathbb{N}$, $\sum_{l \neq j} \frac{\lambda_l}{|\lambda_j - \lambda_j|} \leq Cj \log j$.*

3. *For sufficiently large $j$, we have*

$$\mathsf{E}\left[\sup_{z \in \mathcal{B}_j} \|G_n(z)\|_\infty^2\right] \leq Cn^{-1}(j \log j)^2$$

*and*

$$\mathsf{E}\left[\sup_{z \in \mathcal{B}_j} \|(zI - \Gamma)^{-1/2} X_1\|^2\right] \leq Cj \log j.$$

4. *We have that $\sup_{z \in \mathcal{B}_j} \|K_n(z)\|_\infty \mathbb{I}_{\mathcal{E}_j} \leq C$ almost surely and $\mathsf{P}(\mathcal{E}_j^c) \leq Cn^{-1/2} j \log j$.*

Write event $\mathcal{A}_n = \{\forall j \in \{1, \ldots, h_n\}, |\hat{\lambda}_j - \lambda_j| < \delta_j/2\}$ for each $n \in \mathbb{N}$. The Lemma 5 of [CMS] explains that the random contours $\hat{\mathcal{B}}_j$ in the integrals can be replaced by the population counterparts $\mathcal{B}_j$ asymptotically based on the asymptotic ignorability of $\mathsf{P}(\mathcal{A}_n)$. However, we technically refine this result to incorporate a certain approximation error omitted in previous proofs (cf. page 344 of [CMS]). Specifically, [CMS] approximated $|\hat{\lambda}_j - \lambda|$ by $|\langle(\hat{\Gamma}_n - \Gamma)e_j, e_j\rangle|$ while ignoring the approximation error $|\hat{\lambda}_j - \lambda_j - \langle(\hat{\Gamma}_n - \Gamma)e_j, e_j\rangle|$. However, this approximation error may not be negligible, which requires an additional condition related to both the truncation parameter $h_n$ and the eigengaps $\{\delta_j\}$. Hence, we state the following lemma separately and provide its proof.

**Lemma 2.**

1. *It holds that*

$$\hat{\Pi}_{h_n} - \Pi_{h_n} = \frac{1}{2\pi\iota}\int_{\mathcal{C}_n}\{(zI - \hat{\Gamma}_n)^{-1} - (zI - \Gamma)^{-1}\}dz + r_{1n}\mathbb{I}_{\mathcal{A}_n^c},$$
$$\hat{\Gamma}^{\dagger}_{h_n} - \Gamma^{\dagger}_{h_n} = \frac{1}{2\pi\iota}\int_{\mathcal{C}_n} z^{-1}\{(zI - \hat{\Gamma}_n)^{-1} - (zI - \Gamma)^{-1}\}dz + r_{2n}\mathbb{I}_{\mathcal{A}_n^c},$$

*where*

$$r_{1n} = \hat{\Pi}_{h_n} - \frac{1}{2\pi\iota}\int_{\mathcal{C}_n}(zI - \hat{\Gamma}_n)^{-1}dz,$$
$$r_{2n} = \hat{\Gamma}^{\dagger}_{h_n} - \frac{1}{2\pi\iota}\int_{\mathcal{C}_n} z^{-1}(zI - \hat{\Gamma}_n)^{-1}dz.$$

2. *Suppose that Conditions (A2) and (A4) in the main paper hold. We then have that*

$$\mathsf{P}(\mathcal{A}_n^c) \leq C_1 n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2} + C_2 n^{-1/2}\sum_{j=1}^{h_n} j\log j.$$

*Proof.* On $\mathcal{A}_n$, since $\hat{\lambda}_j$ lies in $\mathcal{B}_j$, we have

$$\hat{\Pi}_{h_n} = \frac{1}{2\pi\iota}\int_{\hat{\mathcal{C}}_{h_n}}(zI - \hat{\Gamma}_n)^{-1}dz$$
$$= \frac{1}{2\pi\iota}\int_{\mathcal{C}_n}(zI - \hat{\Gamma}_n)^{-1}dz.$$

This implies that

$$\left\{\hat{\Pi}_{h_n} - \frac{1}{2\pi\iota}\int_{\mathcal{C}_n}(zI - \hat{\Gamma}_n)^{-1}dz\right\}\mathbb{I}_{\mathcal{A}_n} = 0,$$

and hence,

$$\hat{\Pi}_{h_n} - \Pi_{h_n} = \frac{1}{2\pi\iota}\int_{\mathcal{C}_n}\{(zI - \hat{\Gamma}_n)^{-1} - (zI - \Gamma)^{-1}\}dz + r_{1n}\mathbb{I}_{\mathcal{A}_n^c},$$

where $r_{1n} = \hat{\Pi}_{h_n} - \frac{1}{2\pi\iota}\int_{\mathcal{C}_n}(zI - \hat{\Gamma}_n)^{-1}dz$. One can derive the second equality for $\hat{\Gamma}^{\dagger}_{h_n}$ with the remainder term $r_{2n} = \hat{\Gamma}^{\dagger}_{h_n} - \frac{1}{2\pi\iota}\int_{\mathcal{C}_n} z^{-1}(zI - \hat{\Gamma}_n)^{-1}dz$ in the same way.

For the second part, We first claim that

$$|\hat{\lambda}_j - \lambda_j - \langle(\hat{\Gamma}_n - \Gamma)e_j, e_j\rangle| \leq C\delta_j^{-1}\|\hat{\Gamma}_n - \Gamma\|_\infty^2. \quad (2.18)$$

To see this, set $\hat{c}_j = \mathrm{sign}(\langle \hat{e}_j, e_j\rangle)$. Note that

$$\begin{aligned}\langle (\hat{\Gamma}_n - \Gamma)\hat{e}_j, \hat{c}_j e_j\rangle &= \langle \hat{\Gamma}_n \hat{e}_j, \hat{c}_j e_j\rangle - \langle \hat{e}_j, \hat{c}_j \Gamma e_j\rangle = \langle \hat{\lambda}_j \hat{e}_j, \hat{c}_j e_j\rangle - \langle \hat{e}_j, \hat{c}_j \lambda_j e_j\rangle \\ &= (\hat{\lambda}_j - \lambda_j)\langle \hat{e}_j, \hat{c}_j e_j\rangle = (\hat{\lambda}_j - \lambda_j)(\langle \hat{e}_j, \hat{c}_j e_j\rangle - 1) + (\hat{\lambda}_j - \lambda_j) \\ &= (\hat{\lambda}_j - \lambda_j)\langle \hat{e}_j, \hat{c}_j e_j - \hat{e}_j\rangle + (\hat{\lambda}_j - \lambda_j),\end{aligned}$$

which implies that

$$|\hat{\lambda}_j - \lambda_j - \langle (\hat{\Gamma}_n - \Gamma)\hat{e}_j, \hat{c}_j e_j\rangle| = |\hat{\lambda}_j - \lambda_j||\langle \hat{e}_j, \hat{e}_j - \hat{c}_j e_j\rangle| \le |\hat{\lambda}_j - \lambda_j|\|\hat{e}_j - \hat{c}_j e_j\|$$

On the other hand, we have

$$\begin{aligned}&|\langle (\hat{\Gamma}_n - \Gamma)e_j, e_j\rangle - \langle (\hat{\Gamma}_n - \Gamma)\hat{e}_j, \hat{c}_j e_j\rangle| \\ =&|\langle (\hat{\Gamma}_n - \Gamma)e_j, e_j - \hat{c}_j \hat{e}_j\rangle + \langle (\hat{\Gamma}_n - \Gamma)e_j, \hat{c}_j \hat{e}_j\rangle - \langle (\hat{\Gamma}_n - \Gamma)\hat{e}_j, \hat{c}_j e_j\rangle| \\ =&|\langle (\hat{\Gamma}_n - \Gamma)e_j, \hat{c}_j e_j - \hat{e}_j\rangle| \le \|\hat{\Gamma}_n - \Gamma\|_\infty \|\hat{c}_j e_j - \hat{e}_j\|\end{aligned}$$

Combining these two results, by Lemmas 2.2 and 2.3 of [20], we have

$$\begin{aligned}&|\hat{\lambda}_j - \lambda_j - \langle (\hat{\Gamma}_n - \Gamma)e_j, e_j\rangle| \\ \le&|\hat{\lambda}_j - \lambda_j - \langle (\hat{\Gamma}_n - \Gamma)\hat{e}_j, \hat{c}_j e_j\rangle| + |\langle (\hat{\Gamma}_n - \Gamma)\hat{e}_j, \hat{c}_j e_j\rangle - \langle (\hat{\Gamma}_n - \Gamma)e_j, e_j\rangle| \\ \le&|\hat{\lambda}_j - \lambda_j|\|\hat{e}_j - \hat{c}_j e_j\| + \|\hat{\Gamma}_n - \Gamma\|_\infty \|\hat{c}_j e_j - \hat{e}_j\| \le C\delta_j^{-1}\|\hat{\Gamma}_n - \Gamma\|_\infty^2,\end{aligned}$$

which verifies the inequality (2.18).

Note that

$$\mathsf{P}(\mathcal{A}_n^c) \le \sum_{j=1}^{h_n} \mathsf{P}(|\hat{\lambda}_j - \lambda_j| \ge \delta_j/2) \le 2\sum_{j=1}^{h_n} \delta_j^{-1}\mathsf{E}[|\hat{\lambda}_j - \lambda_j|]$$

by Markov inequality. We also have $\mathsf{E}[\langle(\hat{\Gamma}_n-\Gamma)e_j,e_j\rangle^2]\le Cn^{-1}\lambda_j^2$ (cf. page 341 of [CMS]). Then, by the inequality (2.18) derived above, Theorem 2.5 of [20], and Lemma 1, we have that

$$\begin{aligned}\mathsf{P}(\mathcal{A}_n^c) &\le C_1\sum_{j=1}^{h_n}\delta_j^{-2}\mathsf{E}[\|\hat{\Gamma}_n-\Gamma\|_\infty^2]+C_2\sum_{j=1}^{h_n}\delta_j^{-1}\mathsf{E}[|\langle(\hat{\Gamma}_n-\Gamma)e_j,e_j\rangle|]\\ &\le C_1n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}+C_2n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1}\lambda_j\\ &\le C_1n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}+C_2n^{-1/2}\sum_{j=1}^{h_n}(j+1)\\ &\le C_1n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}+C_2n^{-1/2}\sum_{j=1}^{h_n}j\log j.\end{aligned}$$

□

*Remark* 6. From the following section, we investiage the rates of convergence of different quantities based on perturbation theory. In the proofs, the quantities related to either $\mathcal{E}_j^c$ or $\mathcal{A}_n^c$ will be negligible due to the following arguments.

1. Let $Q_j$ be any non-negative quantity (that can be either random or fixed and can depend on $n$ or not). Note that $\mathbb{I}_{\mathcal{E}_j^c}=0$ implies that $Q_j\mathbb{I}_{\mathcal{E}_j^c}=0$. Let $\eta>0$ be given. If $\sum_{j=1}^{h_n}Q_j\mathbb{I}_{\mathcal{E}_j^c}>\eta$, then $\sum_{j=1}^{h_n}Q_j\mathbb{I}_{\mathcal{E}_j^c}\ne 0$, and hence, there exists $j$ such that $\mathbb{I}_{\mathcal{E}_j^c}\ne 0$.

   (a) Suppose either $\tilde{\mathsf{P}}=\mathsf{P}$ or $\tilde{\mathsf{P}}=\mathsf{P}(\cdot|X_0)$. We then see that

   $$\begin{aligned}\tilde{\mathsf{P}}\left(\sum_{j=1}^{h_n}Q_j\mathbb{I}_{\mathcal{E}_j^c}>\eta\right) &\le \sum_{j=1}^{h_n}\tilde{\mathsf{P}}(\mathbb{I}_{\mathcal{E}_j^c}\ne 0)=\sum_{j=1}^{h_n}\tilde{\mathsf{P}}(\mathcal{E}_j^c)=\sum_{j=1}^{h_n}\mathsf{P}(\mathcal{E}_j^c)\\ &\le Cn^{-1/2}\sum_{j=1}^{h_n}j\log j\end{aligned}$$

   by Lemma 1.

   (b) Suppose either $\tilde{\mathsf{P}}=\mathsf{P}(\cdot|\mathcal{X}_n)$ or $\tilde{\mathsf{P}}=\mathsf{P}(\cdot|\mathcal{X}_n,X_0)$. We then see that

   $$\tilde{\mathsf{P}}\left(\sum_{j=1}^{h_n}Q_j\mathbb{I}_{\mathcal{E}_j^c}>\eta\right)\le \sum_{j=1}^{h_n}\tilde{\mathsf{P}}(\mathbb{I}_{\mathcal{E}_j^c}\ne 0)=\sum_{j=1}^{h_n}\tilde{\mathsf{P}}(\mathcal{E}_j^c)=\sum_{j=1}^{h_n}\tilde{\mathsf{E}}[\mathbb{I}_{\mathcal{E}_j^c}]=\sum_{j=1}^{h_n}\mathbb{I}_{\mathcal{E}_j^c}$$

and for each $\eta' > 0$,

$$\mathsf{P}\left(\sum_{j=1}^{h_n} \mathbb{I}_{\mathcal{E}_j^c} > \eta'\right) \leq \sum_{j=1}^{h_n} \mathsf{P}(\mathbb{I}_{\mathcal{E}_j^c} \neq 0) = \sum_{j=1}^{h_n} \mathsf{P}(\mathcal{E}_j^c) \leq Cn^{-1/2} \sum_{j=1}^{h_n} j \log j$$

by Lemma 1.

Thus, any quantities related to $\mathbb{I}_{\mathcal{E}_j^c}$ (or their sums) are asymptotically negligible or ignorable under any choice of conditional probabilities $\tilde{\mathsf{P}}$ if $n^{-1/2} \sum_{j=1}^{h_n} j \log j \to 0$ as $n \to \infty$. This helps to theoretically guarantee that $\sup_{z \in \mathcal{B}_j} \|K_n(z)\|_\infty$ is bounded above almost surely (with upper bound not depending on $j$) based on Lemma 1.

2. Let $Q_n$ be any non-negative quantity (that can be either random or fixed and can depend on $n$ or not). Note that $\mathbb{I}_{\mathcal{A}_n^c} = 0$ implies that $Q_n \mathbb{I}_{\mathcal{A}_n^c} = 0$. Let $\eta > 0$ be given. If $Q_n \mathbb{I}_{\mathcal{A}_n^c} > \eta$, then $Q_n \mathbb{I}_{\mathcal{A}_n^c} \neq 0$, and hence, $\mathbb{I}_{\mathcal{A}_n^c} \neq 0$.

   (a) Suppose either $\tilde{\mathsf{P}} = \mathsf{P}$ or $\tilde{\mathsf{P}} = \mathsf{P}(\cdot|X_0)$. We then see that

$$\begin{aligned}\tilde{\mathsf{P}}(Q_n \mathbb{I}_{\mathcal{A}_n^c} > \eta) &\leq \tilde{\mathsf{P}}(Q_n \mathbb{I}_{\mathcal{A}_n^c} \neq 0) \leq \tilde{\mathsf{P}}(\mathbb{I}_{\mathcal{A}_n^c} \neq 0) = \tilde{\mathsf{P}}(\mathcal{A}_n^c) = \mathsf{P}(\mathcal{A}_n^c) \\ &\leq C_1 n^{-1} \sum_{j=1}^{h_n} \delta_j^{-2} + C_2 n^{-1/2} \sum_{j=1}^{h_n} j \log j.\end{aligned}$$

   (b) Suppose either $\tilde{\mathsf{P}} = \mathsf{P}(\cdot|\mathcal{X}_n)$ or $\tilde{\mathsf{P}} = \mathsf{P}(\cdot|\mathcal{X}_n, X_0)$. We then see that

$$\tilde{\mathsf{P}}(Q_n \mathbb{I}_{\mathcal{A}_n^c} > \eta) \leq \tilde{\mathsf{P}}(Q_n \mathbb{I}_{\mathcal{A}_n^c} \neq 0) \leq \tilde{\mathsf{P}}(\mathbb{I}_{\mathcal{A}_n^c} \neq 0) = \tilde{\mathsf{P}}(\mathcal{A}_n^c) = \tilde{\mathsf{E}}[\mathbb{I}_{\mathcal{A}_n^c}] = \mathbb{I}_{\mathcal{A}_n^c},$$

   and for each $\eta' > 0$,

$$\mathsf{P}(\mathbb{I}_{\mathcal{A}_n^c} > \eta') \leq \mathsf{P}(\mathbb{I}_{\mathcal{A}_n^c} \neq 0) = \mathsf{P}(\mathcal{A}_n^c) \leq C_1 n^{-1} \sum_{j=1}^{h_n} \delta_j^{-2} + C_2 n^{-1/2} \sum_{j=1}^{h_n} j \log j.$$

   Thus, due to Lemma 17, any quantities related to $\mathbb{I}_{\mathcal{A}_n^c}$ are also asymptotically ignorable under any choice of conditional probabilities $\tilde{\mathsf{P}}$ if $n^{-1} \sum_{j=1}^{h_n} \delta_j^{-2} \to 0$ and $n^{-1/2} \sum_{j=1}^{h_n} j \log j \to 0$ as $n \to \infty$. This aspect theoretically guarantees that the random contour $\hat{\mathcal{C}}_{h_n}$ for $\hat{\Pi}_{h_n}$ and $\hat{\Gamma}_{h_n}^{\dagger}$ can be replaced with the fixed contour $\mathcal{C}_n$.

In what follows, we suppose that Conditions from (A1) to (A6) and $h_n t_{h_n}(X_0)^{-1} = O_{\mathsf{P}}(1)$ hold unless otherwise stated.

### 2.9.1 Preliminaries

Before presenting the main theory, the following lemmas concern the differences $\hat{\Pi}_{h_n} - \Pi_{h_n}$ and $\hat{\Gamma}^{\dagger}_{h_n} - \Gamma^{\dagger}_{h_n}$ between the sample and population operators and are introduced due to frequent usage. Most of these follow from similar arguments as in Proposition 3 in [CMS], but all the proofs will be provided. For conditional arguments, the lemmas are provded either conditionally or unconditionally given on $\mathcal{X}_n$ and $X_0$. In what follows, $\mathsf{E}^X[\cdot] = \mathsf{E}[\cdot|\mathcal{X}_n, X_0]$ and $\mathsf{P}^X(\cdot) = \mathsf{P}(\cdot|\mathcal{X}_n, X_0)$ denotes the conditional expectation and conditional probability, respectively, given $\mathcal{X}_n$ and $X_0$.

**Lemma 3.** *As $n \to \infty$, if $n^{-1/2}\sum_{j=1}^{h_n} j\log j \to 0$, we have the following.*

1. $\|\hat{\Pi}_{h_n} - \Pi_{h_n}\|_\infty = O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n} j\log j\right)$.
2. $\|\hat{\Gamma}^{\dagger}_{h_n} - \Gamma^{\dagger}_{h_n}\|_\infty = O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n} \delta_j^{-1/2}(j\log j)^{3/2}\right)$.
3. $\|(\hat{\Pi}_{h_n} - \Pi_{h_n})X_0\| = O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n} j\log j\right)$.
4. $\|(\hat{\Gamma}^{\dagger}_{h_n} - \Gamma^{\dagger}_{h_n})X_0\| = O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n} \delta_j^{-1/2}(j\log j)^{3/2}\right)$.

*Proof.* Only the last part is proved, as the remaining parts are similar. We observe that

$$\begin{aligned}
\hat{\Gamma}^{\dagger}_{h_n} - \Gamma^{\dagger}_{h_n} &= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}\left\{(zI-\hat{\Gamma}_n)^{-1} - (zI-\Gamma)^{-1}\right\}dz + r_{2n}\mathbb{I}_{\mathcal{A}_n} \\
&= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}(zI-\hat{\Gamma}_n)^{-1}(\hat{\Gamma}_n - \Gamma)(zI-\Gamma)^{-1}dz + r_{2n}\mathbb{I}_{\mathcal{A}_n} \\
&= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}(zI-\Gamma)^{-1/2}K_n(z)G_n(z)(zI-\Gamma)^{-1/2}dz + r_{2n}\mathbb{I}_{\mathcal{A}_n}.
\end{aligned}$$

This implies that $\|(\hat{\Gamma}^{\dagger}_{h_n} - \Gamma^{\dagger}_{h_n})X_0\| \le C\sum_{j=1}^{h_n} A_j + \|r_{2n}X_0\|\mathbb{I}_{\mathcal{A}_n}$ where

$$A_j = \int_{\mathcal{B}_j} |z|^{-1}\|(zI-\Gamma)^{-1/2}\|_\infty\|K_n(z)\|_\infty\|G_n(z)\|_\infty\|(zI-\Gamma)^{-1/2}X_0\|dz.$$

Note that for all $z \in \mathcal{B}_j$, $|z| \ge \lambda_j - \delta_j/2 \ge \lambda_j/2$. By the equation (5.3) of [21], for $z \in \mathcal{B}_j$, we have

$$\|(zI-\Gamma)^{-1/2}\|_\infty = \left(\min_{l\in\mathbb{N}}|z-\lambda_l|^{1/2}\right)^{-1} = |z-\lambda_j|^{-1/2} = (\delta_j/2)^{-1/2}.$$

Thus, by Lemma 1, we have

$$\begin{aligned}\mathsf{E}[A_j\mathbb{I}_{\mathcal{E}_j}] &= \int_{\mathcal{B}_j} \lambda_j^{-1}\|zI-\Gamma\|_\infty^{-1/2}\mathsf{E}[\|K_n(z)\|_\infty\mathbb{I}_{\mathcal{E}_j}\|G_n(z)\|_\infty]\mathsf{E}[\|(zI-\Gamma)^{-1/2}X_0\|]dz \\ &\le C\int_{\mathcal{B}_j} \lambda_j^{-1}\delta_j^{-1/2}\mathsf{E}[\|G_n(z)\|_\infty]\mathsf{E}[\|(zI-\Gamma)^{-1/2}X_0\|]dz \\ &\le C\lambda_j^{-1}\delta_j^{1/2}(n^{-1/2}j\log j)(j\log j)^{1/2} \le Cn^{-1/2}\delta_j^{-1/2}(j\log j)^{3/2},\end{aligned}$$

and hence,

$$\mathsf{E}\left[\sum_{j=1}^{h_n} A_j\mathbb{I}_{\mathcal{E}_j}\right] = O\left(n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right).$$

Consequently, we have the desired result by following the argument in Remark 6 on the reaminder terms related to $\mathcal{E}_j^c$ and $\mathcal{A}_n^c$. □

**Lemma 4.** *As $n\to\infty$, if $n^{-1/2}\sum_{j=1}^{h_n} j\log j \xrightarrow[n\to\infty]{} 0$, we have the following.*

1. $\mathsf{E}^X[\|(\hat{\Gamma}_{h_n}^\dagger-\Gamma_{h_n}^\dagger)U_n\|] = O_\mathsf{P}\left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)$*. Hence, as $n\to\infty$, if*

$$n^{-1}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\to 0,$$

*then for each $\eta>0$, we have*

$$\mathsf{P}^X(\|(\hat{\Gamma}_{h_n}^\dagger-\Gamma_{h_n}^\dagger)U_n\|>\eta)\xrightarrow{\mathsf{P}} 0.$$

2. $\|(\hat{\Gamma}_{h_n}^\dagger-\Gamma_{h_n}^\dagger)U_n\| = O_\mathsf{P}\left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)$.

*Proof.* We observe that

$$\begin{aligned}\hat{\Gamma}_{h_n}^\dagger-\Gamma_{h_n}^\dagger &= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{B_j} z^{-1}\left\{(zI-\hat{\Gamma}_n)^{-1}-(zI-\Gamma)^{-1}\right\}dz + r_{2n}\mathbb{I}_{\mathcal{A}_n} \\ &= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{B_j} z^{-1}(zI-\hat{\Gamma}_n)^{-1}(\hat{\Gamma}_n-\Gamma)(zI-\Gamma)^{-1}dz + r_{2n}\mathbb{I}_{\mathcal{A}_n} \\ &= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{B_j} z^{-1}(zI-\Gamma)^{-1/2}K_n(z)G_n(z)(zI-\Gamma)^{-1/2}dz + r_{2n}\mathbb{I}_{\mathcal{A}_n}.\end{aligned}$$

This implies that $\|(\hat{\Gamma}_{h_n}^\dagger-\Gamma_{h_n}^\dagger)U_n\| \le C\sum_{j=1}^{h_n} A_j + \|r_{2n}U_n\|\mathbb{I}_{\mathcal{A}_n}$ where

$$A_j = \int_{\mathcal{B}_j}\frac{1}{|z|}\|(zI-\Gamma)^{-1/2}\|_\infty\|K_n(z)\|_\infty\|G_n(z)\|_\infty\|(zI-\Gamma)^{-1/2}U_n\|dz.$$

1. Notice that

$$\mathsf{E}^X\left[\|(zI-\Gamma)^{-1/2}U_n\|^2\right]=\sigma_\varepsilon^2 n^{-2}\sum_{i=1}^n\|(zI-\Gamma)^{-1/2}X_i\|^2,$$

and

$$\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\mathsf{E}^X\left[\|(zI-\Gamma)^{-1/2}U_n\|^2\right]\right]=\sigma_\varepsilon^2 n^{-1}\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_1\|^2\right]\leq Cn^{-1}j\log j$$

by Lemma 1. Note that for all $z\in\mathcal{B}_j$, $|z|\geq\lambda_j-\delta_j/2\geq\lambda_j/2$. By the equation (5.3) of [21], for $z\in\mathcal{B}_j$, we have

$$\|(zI-\Gamma)^{-1/2}\|_\infty=\left(\min_{l\in\mathbb{N}}|z-\lambda_l|^{1/2}\right)^{-1}=|z-\lambda_j|^{-1/2}=(\delta_j/2)^{-1/2}.$$

Thus, we have

$$\begin{aligned}\sum_{j=1}^{h_n}\mathsf{E}^X[A_j]\mathbb{I}_{\mathcal{E}_j}\leq C\sum_{j=1}^{h_n}&\operatorname{diam}(\mathcal{B}_j)\delta_j^{-1}\delta_j^{-1/2}\sup_{z\in\mathcal{B}_j}\|K_n(z)\|_\infty\mathbb{I}_{\mathcal{E}_j}\\&\times\sup_{z\in\mathcal{B}_j}\|G_n(z)\|_\infty\sup_{z\in\mathcal{B}_j}\mathsf{E}^X\left[\|(zI-\Gamma)^{-1/2}U_n\|\right]\\\leq C\sum_{j=1}^{h_n}&\delta_j^{-1/2}\sup_{z\in\mathcal{B}_j}\|G_n(z)\|_\infty\sup_{z\in\mathcal{B}_j}\mathsf{E}^X\left[\|(zI-\Gamma)^{-1/2}U_n\|\right],\end{aligned}$$

and

$$\begin{aligned}\mathsf{E}\left[\sum_{j=1}^{h_n}\mathsf{E}^X[A_j]\mathbb{I}_{\mathcal{E}_j}\right]&\leq C\sum_{j=1}^{h_n}\delta_j^{-1/2}\sqrt{\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|G_n(z)\|_\infty^2\right]}\sqrt{\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\mathsf{E}^X\left[\|(zI-\Gamma)^{-1/2}U_n\|\right]^2\right]}\\&\leq C\sum_{j=1}^{h_n}\delta_j^{-1/2}n^{-1/2}j\log j n^{-1/2}(j\log j)^{1/2}=Cn^{-1}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}.\end{aligned}$$

This entails that as $n\to\infty$, if $n^{-1}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\to 0$, then $\sum_{j=1}^{h_n}\mathsf{E}^X[A_j]\mathbb{I}_{\mathcal{E}_j}\xrightarrow{\mathsf{P}}0$.

2. A similar argument applies here. Note that for all $z\in\mathcal{B}_j$, $|z|\geq\lambda_j-\delta_j/2\geq\lambda_j/2$. By the equation (5.3) of [21], for $z\in\mathcal{B}_j$, we have

$$\|(zI-\Gamma)^{-1/2}\|_\infty=\left(\min_{l\in\mathbb{N}}|z-\lambda_l|^{1/2}\right)^{-1}=|z-\lambda_j|^{-1/2}=(\delta_j/2)^{-1/2}.$$

Thus, we have

$$\begin{aligned}
\mathsf{E}[A_j \mathbb{I}_{\mathcal{E}_j}] &= \int_{\mathcal{B}_j} |z|^{-1} \|(zI-\Gamma)^{-1/2}\|_\infty \mathsf{E}[\|K_n(z)\|_\infty \mathbb{I}_{\mathcal{E}_j} \|G_n(z)\|_\infty \|(zI-\Gamma)^{-1/2} U_n\|] dz \\
&\le C \int_{\mathcal{B}_j} \lambda_j^{-1} \delta_j^{-1/2} \mathsf{E}[\|G_n(z)\|_\infty \|(zI-\Gamma)^{-1/2} U_n\|] dz \\
&\le C \int_{\mathcal{B}_j} \lambda_j^{-1} \delta_j^{-1/2} \sqrt{\mathsf{E}[\|G_n(z)\|_\infty^2]} \sqrt{\mathsf{E}[\|(zI-\Gamma)^{-1/2} U_n\|^2]} dz \\
&\le C \delta_j^{-1/2} \frac{\delta_j}{\lambda_j} (n^{-1/2} j \log j) \{n^{-1/2} (j \log j)^{1/2}\} = C n^{-1} \delta_j^{-1/2} (j \log j)^{3/2}.
\end{aligned}$$

We therefore conclude that $\mathsf{E}\left[\sum_{j=1}^{h_n} A_j \mathbb{I}_{\mathcal{E}_j}\right] \le C n^{-1} \sum_{j=1}^{h_n} \delta_j^{-1/2} (j \log j)^{3/2}$, which implies

$$\sum_{j=1}^{h_n} A_j \mathbb{I}_{\mathcal{E}_j} = O_{\mathsf{P}}\left( n^{-1} \sum_{j=1}^{h_n} \delta_j^{-1/2} (j \log j)^{3/2} \right).$$

Consequently, we have the desired result by following the argument in Remark 6 on the reaminder terms related to $\mathcal{E}_j^c$ and $\mathcal{A}_n^c$. □

**Lemma 5.** *We have the following.*

1. $\mathsf{E}^X[\|\hat{\Gamma}_{h_n}^\dagger U_n\|^2] = O_{\mathsf{P}}\left(n^{-1} \sum_{j=1}^{h_n} \lambda_j^{-1}\right)$. *Hence, as $n \to \infty$, if $n^{-1} \sum_{j=1}^{h_n} \lambda_j^{-1} \to 0$, then for any $\eta > 0$,*

$$\mathsf{P}^X(\|\hat{\Gamma}_{h_n}^\dagger U_n\| > \eta) \xrightarrow{\mathsf{P}} 0.$$

2. $\|\Gamma_{h_n}^\dagger U_n\| = O_{\mathsf{P}}\left(n^{-1/2} \sqrt{\sum_{j=1}^{h_n} \lambda_j^{-1}}\right)$.

*Proof.*

1. Note that

$$\begin{aligned}
\mathsf{E}^X[\|\hat{\Gamma}_{h_n}^\dagger U_n\|^2] &= \mathsf{E}^X\left[\left\| n^{-1} \sum_{i=1}^n \hat{\Gamma}_{h_n}^\dagger X_i \varepsilon_i \right\|^2\right] = n^{-2} \sum_{i=1}^n \mathsf{E}^X[\varepsilon_i^2] \|\hat{\Gamma}_{h_n}^\dagger X_i\|^2 \\
&= \sigma_\varepsilon^2 n^{-2} \sum_{i=1}^n \left\| \sum_{j=1}^{h_n} \lambda_j^{-1} \langle X_i, e_j \rangle e_j \right\|^2 = \frac{\sigma_\varepsilon^2}{n^2} \sum_{i=1}^n \sum_{j=1}^{h_n} \lambda_j^{-2} \langle X_i, e_j \rangle^2,
\end{aligned}$$

which implies that

$$\mathsf{E}\left[\mathsf{E}^X[\|\hat{\Gamma}_{h_n}^{\dagger}U_n\|^2]\right] = \frac{\sigma_\varepsilon^2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{h_n}\lambda_j^{-1} = \frac{\sigma_\varepsilon^2}{n}\sum_{j=1}^{h_n}\lambda_j^{-1}.$$

We thus have the desired result.

2. We first note that $\|\Gamma_{h_n}^{\dagger}U_n\|^2 = \left\|\sum_{j=1}^{h_n}\lambda_j^{-1}\langle U_n, e_j\rangle e_j\right\|^2 = \sum_{j=1}^{h_n}\lambda_j^{-2}\langle U_n, e_j\rangle^2$ and

$$\begin{aligned}
\mathsf{E}[\langle U_n, e_j\rangle^2] &= \mathsf{E}\left[\left(n^{-1}\sum_{i=1}^{n}\varepsilon_i\langle X_i, e_j\rangle\right)^2\right] \\
&= n^{-2}\sum_{i=1}^{n}\mathsf{E}[\varepsilon_i^2\langle X_i, e_j\rangle^2] + n^{-2}\sum_{i\neq i'}\mathsf{E}[\varepsilon_i\varepsilon_i'\langle X_i, e_j\rangle\langle X_{i'}, e_j\rangle]. \\
&= n^{-2}\sum_{i=1}^{n}\mathsf{E}[\mathsf{E}[\varepsilon_i^2\langle X_i, e_j\rangle^2|X_i]] + n^{-2}\sum_{i\neq i'}\mathsf{E}[\mathsf{E}[\varepsilon_i\varepsilon_i'\langle X_i, e_j\rangle\langle X_{i'}, e_j\rangle|X_i, X_{i'}]] \\
&= n^{-2}\sum_{i=1}^{n}\mathsf{E}[\langle X_i, e_j\rangle^2\mathsf{E}[\varepsilon_i^2|X_i]] + n^{-2}\sum_{i\neq i'}\mathsf{E}[\langle X_i, e_j\rangle\langle X_{i'}, e_j\rangle\mathsf{E}[\varepsilon_i|X_i]\mathsf{E}[\varepsilon_i'|X_{i'}]] \\
&= n^{-2}\sum_{i=1}^{n}\sigma_\varepsilon^2\mathsf{E}[\langle X_i, e_j\rangle^2] = \frac{\sigma_\varepsilon^2}{n}\mathsf{E}[\langle X_1, e_j\rangle] = \frac{\sigma_\varepsilon^2}{n}\lambda_j.
\end{aligned}$$

This implies that

$$\mathsf{E}\left[\|\Gamma_{h_n}^{\dagger}U_n\|^2\right] = \sum_{j=1}^{h_n}\lambda_j^{-2}\mathsf{E}[\langle U_n, e_j\rangle^2] \leq Cn^{-1}\sum_{j=1}^{h_n}\lambda_j^{-1},$$

and hence, $\|\Gamma_{h_n}^{\dagger}U_n\| = O_{\mathsf{P}}\left(n^{-1/2}\sqrt{\sum_{j=1}^{h_n}\lambda_j^{-1}}\right)$.

□

**Theorem 4** (Consistency of the FPC estimator)**.** *As* $n \to \infty$*, if* $n^{-1/2}h_n^2\log h_n \to 0$*, then we have* $\|\hat{\beta}_{h_n} - \beta\| \xrightarrow{\mathsf{P}} 0$*.*

*Proof.* Note that the remainder terms related to $\mathcal{E}_j^c$ and $\mathcal{A}_n^c$ are negligible by following the argument in Remark 6. Then, by Lemmas 3-5 and the decomposition (3.38), we see that

$$\begin{aligned}\|\hat{\beta}_{h_n}-\beta\| &\le \|(\hat{\Gamma}^{\dagger}_{h_n}-\Gamma^{\dagger}_{h_n})U_n\|+\|\Gamma^{\dagger}_{h_n}U_n\|+\|(\hat{\Pi}_{h_n}-\Pi_{h_n})\beta\|+\|(\Pi_{h_n}-I)\beta\| \\ &= O_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)+O_{\mathsf{P}}\left(\sqrt{n^{-1}\sum_{j=1}^{h_n}\lambda_j^{-1}}\right) \\ &\quad + O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n} j\log j\right)+O\left(\sqrt{\sum_{j>h_n}\langle\beta,e_j\rangle^2}\right).\end{aligned} \tag{2.19}$$

Note from Cauchy-Schwarz inequality that $\left(\sum_{j=1}^{h_n}\delta_j^{-1}\right)^2 \le h_n\sum_{j=1}^{h_n}\delta_j^{-2}$, which implies that

$$\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1}\right)^2 = n^{-1}h_n^{-1}\left(\sum_{j=1}^{h_n}\delta_j^{-1}\right)^2 \le n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}.$$

We also have that $\left(\sum_{j=1}^{h_n}\delta_j^{-1/2}\right)^2 \le h_n\sum_{j=1}^{h_n}\delta_j^{-1}$, which implies that

$$\left(n^{-1/4}h_n^{-3/4}\sum_{j=1}^{h_n}\delta_j^{-1/2}\right)^2 = n^{-1/2}h_n^{-3/2}\left(\sum_{j=1}^{h_n}\delta_j^{-1/2}\right)^2 \le n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1}.$$

Thus, under Condition (A6), as $n\to\infty$, we have $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1}\to 0$ and $n^{-1/4}h_n^{-3/4}\sum_{j=1}^{h_n}\delta_j^{-1/2}\to 0$. The first term in (2.19) is bounded as

$$\begin{aligned}&n^{-1}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2} \\ \le& n^{-1}h_n^{3/2}(\log h_n)^{3/2}\sum_{j=1}^{h_n}\delta_j^{-1/2} = \left(n^{-1/4}h_n^{-3/4}\sum_{j=1}^{h_n}\delta_j^{-1/2}\right)\{n^{-3/4}h_n^{9/4}(\log h_n)^{3/2}\} \\ =& o(1)\left\{\frac{h_n^3(\log h_n)^2}{n}\right\}^{3/4}.\end{aligned}$$

Next, the second term in (2.19) is bounded as

$$n^{-1}\sum_{j=1}^{h_n}\lambda_j^{-1} \le \left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1}\right)(n^{-1/2}h_n^{1/2}) = o(1)\left(\frac{h_n}{n}\right)^{1/2}.$$

Finally, the third term in (2.19) is bounded as

$$n^{-1/2}\sum_{j=1}^{h_n} j\log j \le n^{-1/2}h_n^2\log h_n = \left\{\frac{h_n^4(\log h_n)^2}{n}\right\}^{1/2}.$$

Thus, as $n \to \infty$, if $n^{-1/2}h_n^2 \log h_n \to 0$, we have that $\|\hat{\beta}_{h_n} - \beta\| \xrightarrow{\mathsf{P}} 0$. □

**Corollary 5.** *As $n \to \infty$, if $\hat{\beta}_{h_n} \xrightarrow{\mathsf{P}} \beta$, then we have that $\hat{\sigma}_\varepsilon^2 \equiv n^{-1}\sum_{i=1}^n (Y_i - \langle \hat{\beta}_{h_n}, X_i\rangle)^2 \xrightarrow{\mathsf{P}} \sigma_\varepsilon^2$.*

*Proof.* We may expand

$$\begin{aligned}\hat{\sigma}_\varepsilon^2 &= n^{-1}\sum_{i=1}^n (Y_i - \langle \hat{\beta}_{h_n}, X_i\rangle)^2 = n^{-1}\sum_{i=1}^n (\langle \beta - \hat{\beta}_{h_n}, X_i\rangle + \varepsilon_i)^2 \\ &= n^{-1}\sum_{i=1}^n (\langle \beta - \hat{\beta}_{h_n}, X_i\rangle)^2 + 2n^{-1}\sum_{i=1}^n \varepsilon_i(\langle \beta - \hat{\beta}_{h_n}, X_i\rangle) + n^{-1}\sum_{i=1}^n \varepsilon_i^2. \end{aligned} \tag{2.20}$$

By using the weak law of large numbers, the first and second terms in (2.20) converge to zero in probability as

$$\left| n^{-1}\sum_{i=1}^n (\langle \beta - \hat{\beta}_{h_n}, X_i\rangle)^2 \right| \le n^{-1}\sum_{i=1}^n \|X_i\|^2 \|\hat{\beta}_{h_n} - \beta\|^2 = \{\mathsf{E}[X_1^2] + o_{\mathsf{P}}(1)\} o_{\mathsf{P}}(1) = o_{\mathsf{P}}(1)$$

and

$$\left| n^{-1}\sum_{i=1}^n \varepsilon_i(\langle \beta - \hat{\beta}_{h_n}, X_i\rangle) \right| \le n^{-1}\sum_{i=1}^n |\varepsilon_i| \|X_i\| \|\hat{\beta}_{h_n} - \beta\| = \{\mathsf{E}[|\varepsilon_1| \|X_1\|] + o_{\mathsf{P}}(1)\} o_{\mathsf{P}}(1) = o_{\mathsf{P}}(1),$$

respectively. Thus, since $n^{-1}\sum_{i=1}^n \varepsilon_i^2 \xrightarrow{\mathsf{P}} \sigma_\varepsilon^2$, we have $\hat{\sigma}_\varepsilon^2 \xrightarrow{\mathsf{P}} \sigma_\varepsilon^2$. □

### 2.9.2 The generalized/refined CLT

#### 2.9.2.1 Random bias terms

We first deal with the bias terms in the decomposition (3.38). We call the biases from $(\hat{\Pi}_{h_n} - \Pi_{h_n})\beta$ and $(\hat{\Gamma}_{h_n}^\dagger - \Gamma_{h_n}^\dagger)U_n$ the first and second random bias terms, respectively, and that from $\Pi_{h_n}\beta - \beta$ the non-random bias from now on.

**Proposition 4.** *As $n \to \infty$, if $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}(j \log j)^2 \to 0$, then we have*

$$\sqrt{\frac{n}{t_{h_n}(X_0)}}\left\langle (\hat{\Pi}_{h_n} - \Pi_{h_n})\beta, X_0 \right\rangle \xrightarrow{\mathsf{P}} 0.$$

*Proof.* Note from the Proposition 2 of [CMS] that

$$\begin{aligned}\sqrt{\frac{n}{t_{h_n}(X_0)}}\left\langle (\hat{\Pi}_{h_n} - \Pi_{h_n})\beta, X_0 \right\rangle &= \sqrt{\frac{h_n}{t_{h_n}(X_0)}}\sqrt{\frac{n}{h_n}}\left\langle (\hat{\Pi}_{h_n} - \Pi_{h_n})\beta, X_0 \right\rangle \\ =&\sqrt{\frac{h_n}{t_{h_n}(X_0)}}\left\{ o_P(1) + O_P\left( n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}(j \log j)^2 \right) \right\}\end{aligned}$$

if $n^{-1/2}\sum_{j=1}^{h_n} j\log j \to 0$, as $n\to\infty$. We thus have the desired result. □

The proof of the following proposition about the second random bias cannot be directly obtained from Proposition 3 of [CMS] even though they have the same structure. There are subtle modifications because of conditioning on $\mathcal{X}_n$ and $X_0$.

**Proposition 5.** *As $n\to\infty$, if $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}(j\log j)^2 \to 0$, for any $\eta>0$,*

$$\mathsf{P}^X\left(\left|\sqrt{\frac{n}{t_{h_n}(X_0)}}\left\langle(\hat{\Gamma}^\dagger_{h_n}-\Gamma^\dagger_{h_n})U_n, X_0\right\rangle\right| > \eta\right) \xrightarrow{\mathsf{P}} 0$$

*Proof.* As seen in the proof of Proposition 3 of [CMS], we observe that

$$\begin{aligned}
&\langle(\hat{\Gamma}^\dagger_{h_n}-\Gamma^\dagger_{h_n})U_n, X_0\rangle\\
=&\frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}\left\langle\left\{(zI-\hat{\Gamma}_n)^{-1}-(zI-\Gamma)^{-1}\right\}U_n, X_0\right\rangle dz + \langle r_{2n}U_n, X_0\rangle\mathbb{I}_{\mathcal{A}_n}\\
=&\frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}\left\langle(zI-\hat{\Gamma}_n)^{-1}(\hat{\Gamma}_n-\Gamma)(zI-\Gamma)^{-1}U_n, X_0\right\rangle dz + \langle r_{2n}U_n, X_0\rangle\mathbb{I}_{\mathcal{A}_n}\\
=&\frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}\left\langle(zI-\Gamma)^{-1/2}K_n(z)G_n(z)(zI-\Gamma)^{-1/2}U_n, X_0\right\rangle dz + \langle r_{2n}U_n, X_0\rangle\mathbb{I}_{\mathcal{A}_n}\\
=&\frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}\left\langle K_n(z)G_n(z)(zI-\Gamma)^{-1/2}U_n, (zI-\Gamma)^{-1/2}X_0\right\rangle dz + \langle r_{2n}U_n, X_0\rangle\mathbb{I}_{\mathcal{A}_n}
\end{aligned}$$

This implies that $\left|\langle(\hat{\Gamma}^\dagger_{h_n}-\Gamma^\dagger_{h_n})U_n, X_0\rangle\right| \le C\sum_{j=1}^{h_n} A_j + |\langle r_{2n}U_n, X_0\rangle|\mathbb{I}_{\mathcal{A}_n}$ where

$$A_j = \int_{\mathcal{B}_j} |z|^{-1}\|K_n(z)\|_\infty\|G_n(z)\|_\infty\|(zI-\Gamma)^{-1/2}U_n\|\|(zI-\Gamma)^{-1/2}X_0\|dz.$$

Notice that

$$\mathsf{E}^X\left[\|(zI-\Gamma)^{-1/2}U_n\|^2\right] = \sigma_\varepsilon^2 n^{-2}\sum_{i=1}^n \|(zI-\Gamma)^{-1/2}X_i\|^2,$$

and

$$\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\mathsf{E}^X\left[\|(zI-\Gamma)^{-1/2}U_n\|^2\right]\right] = \sigma_\varepsilon^2 n^{-1}\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_1\|^2\right] \le Cn^{-1}j\log j.$$

This implies that

$$\begin{aligned}
&\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\left\|(zI-\Gamma)^{-1/2}X_0\right\|^2\sup_{z\in\mathcal{B}_j}\mathsf{E}^X\left[\|(zI-\Gamma)^{-1/2}U_n\|^2\right]\right]\\
=&\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\left\|(zI-\Gamma)^{-1/2}X_0\right\|^2\right]\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\mathsf{E}^X\left[\|(zI-\Gamma)^{-1/2}U_n\|^2\right]\right]\\
\leq& Cn^{-1}(j\log j)^2.
\end{aligned}$$

by independence bewteen $X_0$ and $\mathcal{X}_n$. Then, by using the third and fourth parts of Lemma 1,

$$\begin{aligned}
\sum_{j=1}^{h_n}\mathsf{E}^X[A_j]\mathbb{I}_{\mathcal{E}_j}&\leq C\sum_{j=1}^{h_n}\mathrm{diam}(\mathcal{B}_j)\delta_j^{-1}\sup_{z\in\mathcal{B}_j}\|G_n(z)\|_\infty\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|\sup_{z\in\mathcal{B}_j}\mathsf{E}^X\left[\|(zI-\Gamma)^{-1/2}U_n\|\right]\\
&\leq C\sqrt{\sum_{j=1}^{h_n}\sup_{z\in\mathcal{B}_j}\|G_n(z)\|_\infty^2}\sqrt{\sum_{j=1}^{h_n}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|^2\sup_{z\in\mathcal{B}_j}\mathsf{E}^X\left[\|(zI-\Gamma)^{-1/2}U_n\|^2\right]}.
\end{aligned}$$

From Lemma 1, we have $\mathsf{E}\left[\sum_{j=1}^{h_n}\sup_{z\in\mathcal{B}_j}\|G_n(z)\|_\infty^2\right]\leq Cn^{-1}\sum_{j=1}^{h_n}(j\log j)^2$, which implies that

$$\sqrt{\sum_{j=1}^{h_n}\sup_{z\in\mathcal{B}_j}\|G_n(z)\|_\infty^2}=O_{\mathsf{P}}\left(n^{-1/2}\left\{\sum_{j=1}^{h_n}(j\log j)^2\right\}^{1/2}\right).$$

We then bound the remaining term by

$$\mathsf{E}\left[\sum_{j=1}^{h_n}\sup_{z\in\mathcal{B}_j}\left\|(zI-\Gamma)^{-1/2}X_0\right\|^2\sup_{z\in\mathcal{B}_j}\mathsf{E}^X\left[\|(zI-\Gamma)^{-1/2}U_n\|^2\right]\right]\leq Cn^{-1}\sum_{j=1}^{h_n}(j\log j)^2$$

from the independence between $\mathcal{X}_n$ and $X_0$, which implies that

$$\sqrt{\sum_{j=1}^{h_n}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|^2\sup_{z\in\mathcal{B}_j}\mathsf{E}^X\left[\|(zI-\Gamma)^{-1/2}U_n\|^2\right]}=O_{\mathsf{P}}\left(n^{-1/2}\left\{\sum_{j=1}^{h_n}(j\log j)^2\right\}^{1/2}\right).$$

Therefore, we have

$$\mathsf{E}^X\left[\sqrt{\frac{n}{t_{h_n}(X_0)}}\sum_{j=1}^{h_n}A_j\mathbb{I}_{\mathcal{E}_j}\right]=O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}(j\log j)^2\right)$$

and the desired result by following the argument in Remark 6 on the reaminder terms related to $\mathcal{E}_j^c$ and $\mathcal{A}_n^c$. □

In addition, Proposition 18 holds even when $X_0=X_1$ with the help of the following lemma.

**Lemma 6.** *Under the same assumptions of Lemma 1, we have*

$$\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^4\right]\le C(j\log j)^2.$$

*Proof.* Note that

$$\begin{aligned}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^4 &\le 4\left(\sum_{k\neq j}\frac{\lambda_k\xi_k^2}{|\lambda_j-\lambda_k|}+\frac{\lambda_j\xi_j^2}{\delta_j}\right)^2\\ &=4\left\{\sum_{l\neq k,j\neq k}\frac{\lambda_l\lambda_k\xi_l^2\xi_k^2}{|\lambda_j-\lambda_l||\lambda_j-\lambda_k|}+\sum_{k\neq j}\frac{\lambda_j\lambda_k\xi_j^2\xi_k^2}{\delta_j|\lambda_j-\lambda_k|}+\frac{\lambda_j^2\xi_j^4}{\delta_j^2}\right\}.\end{aligned}$$

Due to Condition (A2) and Lemma 1, we have that

$$\begin{aligned}\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^4\right] &\le C\left\{\sum_{l\neq k,j\neq k}\frac{\lambda_l\lambda_k}{|\lambda_j-\lambda_l||\lambda_j-\lambda_k|}+\sum_{k\neq j}\frac{\lambda_j\lambda_k}{\delta_j|\lambda_j-\lambda_k|}+\frac{\lambda_j^2}{\delta_j^2}\right\}\\ &=C\left\{\left(\sum_{k\neq j}\frac{\lambda_k}{|\lambda_j-\lambda_k|}\right)^2+\frac{\lambda_j}{\delta_j}\sum_{k\neq j}\frac{\lambda_k}{|\lambda_j-\lambda_k|}+\frac{\lambda_j^2}{\delta_j^2}\right\}\\ &\le C\left\{(Cj\log j)^2+(j+1)(Cj\log j)+(j+1)^2\right\}\\ &\le C(j\log j)^2.\end{aligned}$$

□

**Proposition 6.** *Proposition 18 holds even when $X_0=X_1$.*

*Proof.* As seen in the proof of Proposition 18, we have $|\langle(\hat{\Gamma}_{h_n}^{\dagger}-\Gamma_{h_n}^{\dagger})U_n,X_1\rangle|\le C\sum_{j=1}^{h_n}A_j+|\langle r_{2n}U_n,X_1\rangle|\mathbb{I}_{\mathcal{A}_n}$ where

$$A_j=\int_{\mathcal{B}_j}|z|^{-1}\|K_n(z)\|_\infty\|G_n(z)\|_\infty\|(zI-\Gamma)^{-1/2}U_n\|\|(zI-\Gamma)^{-1/2}X_1\|dz.$$

By taking the expectation $\mathsf{E}^X$, we have

$$\begin{aligned}\mathsf{E}^X[A_j] &=\int_{\mathcal{B}_j}|z|^{-1}\|K_n(z)\|_\infty\|G_n(z)\|_\infty\mathsf{E}^X[\|(zI-\Gamma)^{-1/2}U_n\|]\|(zI-\Gamma)^{-1/2}X_1\|dz\\ &\le\sigma_\varepsilon\int_{\mathcal{B}_j}|z|^{-1}\|K_n(z)\|_\infty\|G_n(z)\|_\infty\left\{n^{-2}\sum_{i=1}^n\|(zI-\Gamma)^{-1/2}X_i\|^2\right\}^{1/2}\|(zI-\Gamma)^{-1/2}X_1\|dz.\end{aligned}$$

By Lemmas 1 and 6, it holds that

$$\begin{aligned}
&\mathsf{E}\left[\sum_{i=1}^{n}\|(zI-\Gamma)^{-1/2}X_i\|^2\|(zI-\Gamma)^{-1/2}X_1\|^2\right]\\
=&\mathsf{E}[\|(zI-\Gamma)^{-1/2}X_1\|^4]+\sum_{i\neq 1}\mathsf{E}[\|(zI-\Gamma)^{-1/2}X_i\|^2]\mathsf{E}[\|(zI-\Gamma)^{-1/2}X_1\|^2]\\
\leq& Cn(j\log j)^2.
\end{aligned}$$

This implies that

$$\begin{aligned}
&\mathsf{E}\left[\mathsf{E}^X\left[\sum_{j=1}^{h_n}A_j\mathbb{I}_{\mathcal{E}_j}\right]\right]\\
\leq& C\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}|z|^{-1}\mathsf{E}\left[\|G_n(z)\|_\infty\left\{n^{-2}\sum_{i=1}^{n}\|(zI-\Gamma)^{-1/2}X_i\|^2\right\}^{1/2}\|(zI-\Gamma)^{-1/2}X_1\|\right]dz\\
\leq& Cn^{-1}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}|z|^{-1}\mathsf{E}[\|G_n(z)\|_\infty^2]^{1/2}\mathsf{E}\left[\sum_{i=1}^{n}\|(zI-\Gamma)^{-1/2}X_i\|^2\|(zI-\Gamma)^{-1/2}X_1\|^2\right]^{1/2}dz\\
\leq& Cn^{-1}\sum_{j=1}^{h_n}(j\log j)^2
\end{aligned}$$

by a similar argument to the proof of Proposition 18. We finally have the desired result by following the argument of Remark 6. □

#### 2.9.2.2 Variance terms

To prove the weak convergence of the variance term $\sqrt{n/t_{h_n}(X_0)}\langle\Gamma_{h_n}^{\dagger}U_n,X_0\rangle$ in Proposition 2 in the main paper, we need the following lemma.

**Lemma 7.** *We have that* $A_n\equiv\langle(\hat{\Gamma}_n-\Gamma)\Gamma_{h_n}^{\dagger}X_0,\Gamma_{h_n}^{\dagger}X_0\rangle=O_{\mathsf{P}}(n^{-1/2}h_n^2)$.

*Proof.* Note that $A_n = n^{-1}\sum_{i=1}^{n} B_{i,n}$ where

$$\begin{aligned}
B_{i,n} &= \left\langle (X_i \otimes X_i - \Gamma)\sum_{j=1}^{h_n} \lambda_j^{-1}\langle X_0, e_j\rangle e_j, \sum_{l=1}^{h_n} \lambda_l^{-1}\langle X_0, e_l\rangle e_l \right\rangle \\
&= \sum_{j,l}^{h_n} \lambda_j^{-1}\lambda_l^{-1}\langle X_0, e_j\rangle \langle X_0, e_l\rangle \langle (X_i \otimes X_i - \Gamma)e_j, e_l\rangle \\
&= \sum_{j,l}^{h_n} \lambda_j^{-1}\lambda_l^{-1}\langle X_0, e_j\rangle \langle X_0, e_l\rangle (\langle X_i, e_j\rangle \langle X_i, e_l\rangle - \langle \Gamma e_j, e_l\rangle) \\
&= \sum_{j,l}^{h_n} \lambda_j^{-1/2}\lambda_l^{-1/2}\xi_{0j}\xi_{0l}\left(\lambda_j^{1/2}\lambda_l^{1/2}\xi_{ij}\xi_{il} - \lambda_j \mathbb{I}(j=l)\right) \\
&= \sum_{j,l}^{h_n} \xi_{0j}\xi_{0l}\{\xi_{ij}\xi_{il} - \mathbb{I}(j=l)\}
\end{aligned}$$

and $\xi_{ij} = \lambda_j^{-1/2}\langle X_i, e_j\rangle$ so that $\xi_{ij}$ has mean 0 and variance 1, and $\mathsf{E}[\xi_{ij}\xi_{il}] = 0$. Let $J_{i,j,l} = \xi_{0j}\xi_{0l}\{\xi_{ij}\xi_{il} - \mathbb{I}(j=l)\}$. We next establish and bound the expected value of $B_{i,n}$ along three cases presented below. Note that $\mathsf{E}[J_{i,j,l}] = \mathsf{E}[\xi_{0j}^2(\xi_{ij}^2 - 1)] = \mathsf{E}[\xi_{0j}^2]\mathsf{E}[(\xi_{ij}^2 - 1)] = 0$ if $j = l$ and $\mathsf{E}[J_{i,j,l}] = \mathsf{E}[\xi_{0j}\xi_{0l}\xi_{ij}\xi_{il}] = \mathsf{E}[\xi_{0j}\xi_{0l}]\mathsf{E}[\xi_{ij}\xi_{il}] = 0$ if $j \neq l$. This implies that $\mathsf{E}[B_{i,n}] = 0$.

Next, to bound the second moments of $B_{i,n}$, note that $B_{i,n}B_{i',n} = \sum_{j,l,j',l'} J_{i,j,l}J_{i',j',l'}$. we now study the expected values of the products $J_{i,j,l}J_{i',j',l'}$ depending on $(j,l,j',l')$. We consider the first case of $i \neq i'$. We have

$$\mathsf{E}[J_{i,j,l}J_{i',j',l'}] = \mathsf{E}[\xi_{0j}^2(\xi_{ij}^2 - 1)\xi_{0j'}^2(\xi_{i'j'}^2 - 1)] = \mathsf{E}[\xi_{0j}^2\xi_{0j'}^2]\mathsf{E}[(\xi_{ij}^2 - 1)]\mathsf{E}[(\xi_{i'j'}^2 - 1)] = 0$$

if $j = l$ and $j' = l'$,

$$\mathsf{E}[J_{i,j,l}J_{i',j',l'}] = \mathsf{E}[\xi_{0j}^2(\xi_{ij}^2 - 1)\xi_{0j'}\xi_{0l'}\xi_{i'j'}\xi_{i'l'}] = \mathsf{E}[\xi_{0j}^2\xi_{0j'}\xi_{0l'}]\mathsf{E}[(\xi_{ij}^2 - 1)]\mathsf{E}[\xi_{i'j'}\xi_{i'l'}] = 0$$

if $j = l$ and $j' \neq l'$, and

$$\mathsf{E}[J_{i,j,l}J_{i',j',l'}] = \mathsf{E}[\xi_{0j}\xi_{0l}\xi_{ij}\xi_{il}\xi_{0j'}\xi_{0l'}\xi_{ij'}\xi_{il'}] = \mathsf{E}[\xi_{0j}\xi_{0l}\xi_{0j'}\xi_{0l'}]\mathsf{E}[\xi_{ij}\xi_{il}]\mathsf{E}[\xi_{ij'}\xi_{il'}] = 0$$

if $j \neq l$ and $j' \neq l'$. This implies that $\mathsf{E}[B_{i,n}B_{i',n}] = 0$ if $i \neq i'$.

For the other cases with $i = i'$, we can bound $\mathsf{E}[J_{i,j,l}J_{i',j',l'}]$ by using the finite fourth moment assumption $\sup_{j\in\mathbb{N}} \lambda_j^{-2}\mathsf{E}[\langle X, e_j\rangle^4] < \infty$ as below. We have

$$\begin{aligned}
\mathsf{E}[J_{i,j,l}J_{i,j',l'}] &= \mathsf{E}[\xi_{0j}^2(\xi_{ij}^2 - 1)\xi_{0j'}^2(\xi_{ij'}^2 - 1)] = \mathsf{E}[\xi_{0j}^2\xi_{0j'}^2]\mathsf{E}[(\xi_{ij}^2 - 1)(\xi_{ij'}^2 - 1)] \\
&= \mathsf{E}[\xi_{0j}^2\xi_{0j'}^2](\mathsf{E}[\xi_{ij}^2\xi_{ij'}^2] - 1) \\
&\le \sqrt{\mathsf{E}[\xi_{0j}^4]\mathsf{E}[\xi_{0j'}^4]}\sqrt{\mathsf{E}[\xi_{ij}^4]\mathsf{E}[\xi_{ij'}^4]} \le C
\end{aligned}$$

if $j = l$ and $j' = l'$,

$$\begin{aligned}
\mathsf{E}[J_{i,j,l}J_{i,j',l'}] &= \mathsf{E}[\xi_{0j}^2(\xi_{ij}^2 - 1)\xi_{0j'}\xi_{0l'}\xi_{ij'}\xi_{il'}] = \mathsf{E}[\xi_{0j}^2\xi_{0j'}\xi_{0l'}]\mathsf{E}[(\xi_{ij}^2 - 1)\xi_{ij'}\xi_{il'}] \\
&= \mathsf{E}[\xi_{0j}^2\xi_{0j'}\xi_{0l'}]\mathsf{E}[\xi_{ij}^2\xi_{ij'}\xi_{il'}] \\
&\le \sqrt{\mathsf{E}[\xi_{0j}^4]}\sqrt{\mathsf{E}[\xi_{0j'}^2\xi_{0l'}^2]}\sqrt{\mathsf{E}[\xi_{ij}^4]}\sqrt{\mathsf{E}[\xi_{ij'}^2\xi_{il'}^2]} \\
&\le \sqrt{\mathsf{E}[\xi_{0j}^4]}\left(\mathsf{E}[\xi_{0j'}^4]\mathsf{E}[\xi_{0l'}^4]\right)^{1/4}\sqrt{\mathsf{E}[\xi_{ij}^4]}\left(\mathsf{E}[\xi_{ij'}^4]\mathsf{E}[\xi_{il'}^4]\right)^{1/4} \\
&\le C
\end{aligned}$$

if $j = l$ and $j' \ne l'$, and

$$\begin{aligned}
\mathsf{E}[J_{i,j,l}J_{i,j',l'}] &= \mathsf{E}[\xi_{0j}\xi_{0l}\xi_{ij}\xi_{il}\xi_{0j'}\xi_{0l'}\xi_{ij'}\xi_{il'}] = \mathsf{E}[\xi_{0j}\xi_{0l}\xi_{0j'}\xi_{0l'}]\mathsf{E}[\xi_{ij}\xi_{il}\xi_{ij'}\xi_{il'}] \\
&\le \left(\mathsf{E}[\xi_{0j}^4]\mathsf{E}[\xi_{0l}^4]\mathsf{E}[\xi_{0j'}^4]\mathsf{E}[\xi_{0l'}^4]\right)^{1/4}\left(\mathsf{E}[\xi_{ij}^4]\mathsf{E}[\xi_{il}^4]\mathsf{E}[\xi_{ij'}^4]\mathsf{E}[\xi_{il'}^4]\right)^{1/4} \\
&\le C
\end{aligned}$$

if $j \ne l$ and $j' \ne l'$. By combining the bound from above, we find $\mathsf{E}[B_{i,n}^2] = \sum_{j,l,j',l'} \mathsf{E}[J_{i,j,l}J_{i,j',l'}] \le \sum_{j,l,j',l'} C = Ch_n^4$. Since $\mathsf{E}[A_n^2] = n^{-2}\sum_{i=1}^n \mathsf{E}[B_{i,n}^2] \le Cn^{-1}h_n^4$, we hence derive that $A_n = O_{\mathsf{P}}(n^{-1/2}h_n^2)$. □

**Proposition 7.** *Proposition 2 in the main paper also holds even when $X_0 = X_1$.*

*Proof.* Denoting $A_n \equiv \langle(\hat{\Gamma}_n - \Gamma)\Gamma_{h_n}^\dagger X_1, \Gamma_{h_n}^\dagger X_1\rangle$, we have that

$$t_{h_n}(X_1)^{-1}|A_n| \le t_{h_n}(X_1)^{-1}\|\hat{\Gamma}_n - \Gamma\|\|\Gamma_{h_n}^\dagger X_1\|^2 = O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}\lambda_j^{-1}\right)$$

due to $\mathsf{E}[\|\hat{\Gamma}_n - \Gamma\|^2] \le n^{-1}\mathsf{E}[\|X_1\|^4]$ from Theorem 2.5 of [20] and $\mathsf{E}[\|\Gamma_{h_n}^{\dagger} X_1\|^2] = \sum_{j=1}^{h_n} \lambda_j^{-2}\mathsf{E}[\langle X_1, e_j\rangle^2] = \sum_{j=1}^{h_n} \lambda_j^{-1}$. Thus, under Condition (A5), we have that $t_{h_n}(X_1)^{-1}A_n = o_{\mathsf{P}}(1)$ since

$$n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}\lambda_j^{-1} \le n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1} \le \left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}\right)^{1/2} \to 0.$$

The proof of Proposition 2 can be completed with $X_0 = X_1$ by the same argument. □

#### 2.9.2.3 Scaling terms

We now provide the proof for the second part of Theorem 1 in the main paper, which guarantees the interchangeability of $t_{h_n}(X_0)$ and $\hat{t}_{h_n}(X_0)$ in the asymptotics. We exclude the case conditional on both $\mathcal{X}_n$ and $X_0$, as the cases with $\tilde{\mathsf{P}} = \mathsf{P}$ and $\tilde{\mathsf{P}} = \mathsf{P}(\cdot|\mathcal{X}_n, X_0)$ indicate the same result. We re-write the statement for preciseness.

**Proposition 8.** *Suppose that $n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}(j\log j)^2 \to 0$ as $n \to \infty$. The scaling $t_{h_n}(X_0)$ and $\hat{t}_{h_n}(X_0)$ are equivalent in that, for any $\eta > 0$,*

$$\tilde{\mathsf{P}}\left(\left|\frac{\hat{t}_{h_n}(X_0)}{t_{h_n}(X_0)} - 1\right| > \eta\right) \xrightarrow{\mathsf{P}} 0,$$

*where $\tilde{\mathsf{P}}$ denotes one of the conditional probabilities $\mathsf{P}$, $\mathsf{P}(\cdot|\mathcal{X}_n)$, or $\mathsf{P}(\cdot|X_0)$.*

*Proof.* We first observe

$$\hat{t}_{h_n}(X_0) = \langle \hat{\Gamma}_{h_n}^{\dagger} X_0, X_0\rangle = \langle (\hat{\Gamma}_{h_n}^{\dagger} - \Gamma_{h_n}^{\dagger})X_0, X_0\rangle + t_{h_n}(X_0)$$

so that

$$\frac{\hat{t}_{h_n}(X_0)}{t_{h_n}(X_0)} - 1 = t_{h_n}(X_0)^{-1}\langle (\hat{\Gamma}_{h_n}^{\dagger} - \Gamma_{h_n}^{\dagger})X_0, X_0\rangle.$$

To use the perturbation theory, note that

$$\begin{aligned}\hat{\Gamma}_{h_n}^{\dagger} - \Gamma_{h_n}^{\dagger} &= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}\left\{(zI - \hat{\Gamma}_n)^{-1} - (zI - \Gamma)^{-1}\right\}dz + r_{2n}\mathbb{I}_{\mathcal{A}_n^c}\\ &= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}(zI - \Gamma)^{-1/2}K_n(z)G_n(z)(zI - \Gamma)^{-1/2}dz + r_{2n}\mathbb{I}_{\mathcal{A}_n^c}.\end{aligned}$$

This implies that $\|\langle(\hat{\Gamma}_{h_n}^{\dagger} - \Gamma_{h_n}^{\dagger})X_0, X_0\rangle\| \leq C\sum_{j=1}^{h_n} A_j + |\langle r_{2n}X_0, X_0\rangle|\mathbb{I}_{\mathcal{A}_n^c}$, where

$$A_j = \int_{\mathcal{B}_j} |z|^{-1}\|K_n(z)\|_\infty \|G_n(z)\|_\infty \|(zI-\Gamma)^{-1/2}X_0\|^2 dz.$$

Note that for all $z \in \mathcal{B}_j$, $|z| \geq \lambda_j - \delta_j/2 \geq \lambda_j/2$. We now study the convergence rates in probability either conditionally on $X_0$ or $\mathcal{X}_n$, or unconditionally on both by using the third and fourth parts of Lemma 1.

1. Consider the unconditional case. By the third and fourth parts of Lemma 1, we have

$$\begin{aligned}
\mathsf{E}[A_j\mathbb{I}_{\mathcal{E}_j}] &\leq \int_{\mathcal{B}_j} \lambda_j^{-1}\mathsf{E}[\|K_n(z)\|_\infty \mathbb{I}_{\mathcal{E}_j}\|G_n(z)\|_\infty]\mathsf{E}[\|(zI-\Gamma)^{-1/2}X_0\|^2]dz \\
&\leq C\int_{\mathcal{B}_j} \lambda_j^{-1}\mathsf{E}[\|G_n(z)\|_\infty]\mathsf{E}[\|(zI-\Gamma)^{-1/2}X_0\|^2]dz \\
&= C\delta_j\lambda_j^{-1}(n^{-1/2}j\log j)(j\log j) \\
&\leq Cn^{-1/2}(j\log j)^2.
\end{aligned}$$

   This implies that $\mathsf{E}\left[\sum_{j=1}^{h_n} A_j\mathbb{I}_{\mathcal{E}_j}\right] \leq Cn^{-1/2}\sum_{j=1}^{h_n}(j\log j)^2$, and hence,

$$t_{h_n}(X_0)^{-1}\sum_{j=1}^{h_n} A_j\mathbb{I}_{\mathcal{E}_j} = O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n}(j\log j)^2\right).$$

2. Consider the case conditional on $X_0$. By the third and fourth parts of Lemma 1, we have

$$\begin{aligned}
\mathsf{E}^{X_0}[A_j\mathbb{I}_{\mathcal{E}_j}] &\leq \int_{\mathcal{B}_j} \lambda_j^{-1}\mathsf{E}[\|K_n(z)\|_\infty \mathbb{I}_{\mathcal{E}_j}\|G_n(z)\|_\infty]\|(zI-\Gamma)^{-1/2}X_0\|^2 dz \\
&\leq C\int_{\mathcal{B}_j} \lambda_j^{-1}\mathsf{E}[\|G_n(z)\|_\infty]\|(zI-\Gamma)^{-1/2}X_0\|^2 dz \\
&= C\delta_j\lambda_j^{-1}\sup_{z\in\mathcal{B}_j}\mathsf{E}[\|G_n(z)\|_\infty]\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|^2 \\
&\leq C(n^{-1/2}j\log j)\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|^2.
\end{aligned}$$

   This implies that

$$\begin{aligned}
\mathsf{E}\left[\mathsf{E}^{X_0}\left[\sum_{j=1}^{h_n} A_j\mathbb{I}_{\mathcal{E}_j}\right]\right] &\leq Cn^{-1/2}\sum_{j=1}^{h_n} j\log j\,\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|^2\right] \\
&\leq Cn^{-1/2}\sum_{j=1}^{h_n}(j\log j)^2,
\end{aligned}$$

and hence, $\mathsf{E}^{X_0}\left[t_{h_n}(X_0)^{-1}\sum_{j=1}^{h_n} A_j\mathbb{I}_{\mathcal{E}_j}\right] = O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}(j\log j)^2\right)$.

3. Consider the case conditional on $\mathcal{X}_n$. By the third and fourth parts of Lemma 1, we have

$$\begin{aligned}
\mathsf{E}^{\mathcal{X}_n}[A_j\mathbb{I}_{\mathcal{E}_j}] &\leq \int_{\mathcal{B}_j}\lambda_j^{-1}\|K_n(z)\|_\infty\mathbb{I}_{\mathcal{E}_j}\|G_n(z)\|_\infty\mathsf{E}[\|(zI-\Gamma)^{-1/2}X_0\|^2]dz \\
&\leq C(j\log j)\int_{\mathcal{B}_j}\lambda_j^{-1}\|K_n(z)\|_\infty\mathbb{I}_{\mathcal{E}_j}\|G_n(z)\|_\infty dz \\
&= C(j\log j)\sup_{z\in\mathcal{B}_j}\|K_n(z)\|_\infty\mathbb{I}_{\mathcal{E}_j}\|G_n(z)\|_\infty.
\end{aligned}$$

This implies that

$$\begin{aligned}
\mathsf{E}\left[\mathsf{E}^{\mathcal{X}_n}\left[\sum_{j=1}^{h_n}A_j\mathbb{I}_{\mathcal{E}_j}\right]\right] &\leq C\sum_{j=1}^{h_n}(j\log j)\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|K_n(z)\|_\infty\mathbb{I}_{\mathcal{E}_j}\|G_n(z)\|_\infty\right] \\
&\leq C\sum_{j=1}^{h_n}(j\log j)\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|G_n(z)\|_\infty\right] \\
&\leq Cn^{-1/2}\sum_{j=1}^{h_n}(j\log j)^2,
\end{aligned}$$

and hence, $\mathsf{E}^{\mathcal{X}_n}\left[h_n^{-1}\sum_{j=1}^{h_n} A_j\mathbb{I}_{\mathcal{E}_j}\right] = O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}(j\log j)^2\right)$.

We now suppose that $n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}(j\log j)^2 \to 0$ as $n\to\infty$, and let $\eta > 0$ be given. From the fact that

$$\lim_{M\to\infty}\limsup_{n\to\infty}\mathsf{P}^{\mathcal{X}_n}(h_nt_{h_n}(X_0)^{-1} > M) = \lim_{M\to\infty}\limsup_{n\to\infty}\mathsf{P}(h_nt_{h_n}(X_0)^{-1} > M) = 0,$$

we have

$$\begin{aligned}
&\mathsf{P}^{\mathcal{X}_n}\left(t_{h_n}(X_0)^{-1}\sum_{j=1}^{h_n}A_j\mathbb{I}_{\mathcal{E}_j} > \eta\right) \\
\leq&\mathsf{P}^{\mathcal{X}_n}\left(h_nt_{h_n}(X_0)^{-1} > M\right) + \mathsf{P}^{\mathcal{X}_n}\left(h_n^{-1}\sum_{j=1}^{h_n}A_j\mathbb{I}_{\mathcal{E}_j} > \eta/M\right) \\
\leq&\mathsf{P}^{\mathcal{X}_n}\left(h_nt_{h_n}(X_0)^{-1} > M\right) + \frac{M}{\eta}\mathsf{E}^{\mathcal{X}_n}\left[h_n^{-1}\sum_{j=1}^{h_n}A_j\mathbb{I}_{\mathcal{E}_j}\right]
\end{aligned}$$

for each $M > 0$. Let $\{n'\}$ be a subsequence of $\{n\}$. Then, since $\mathsf{E}^{\mathcal{X}_n}\left[h_n^{-1}\sum_{j=1}^{h_n} A_j\mathbb{I}_{\mathcal{E}_j}\right] \xrightarrow{\mathsf{P}} 0$ as $n\to\infty$, there exists a further subsequence $\{n''\}\subseteq\{n'\}$ of $\{n\}$ and an almost sure set

$D \in \mathcal{F}$ (i.e., $\mathsf{P}(D) = 1$) such that $\mathsf{E}^{\mathcal{X}_{n''}}\left[h_{n''}^{-1}\sum_{j=1}^{h_{n''}} A_j \mathbb{I}_{\mathcal{E}_j}\right] \to 0$ on $D$. On $D$, we see that

$$\limsup_{n''\to\infty} \mathsf{P}^{\mathcal{X}_{n''}}\left(t_{h_{n''}}(X_0)^{-1}\sum_{j=1}^{h_{n''}} A_j \mathbb{I}_{\mathcal{E}_j} > \eta\right) \le \limsup_{n''\to\infty} \mathsf{P}^{\mathcal{X}_{n'}}(h_{n''}t_{h_{n''}}(X_0)^{-1} > M) \to 0$$

as $M \to \infty$. Since this holds for each subsequence $\{n'\} \subseteq \{n\}$, as $n \to \infty$, if $n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}(j\log j)^2 \to 0$, we have that

$$\mathsf{P}^{\mathcal{X}_n}\left(t_{h_n}(X_0)^{-1}\sum_{j=1}^{h_n} A_j \mathbb{I}_{\mathcal{E}_j} > \eta\right) \xrightarrow{\mathsf{P}} 0.$$

We finally have the desired result by following the argument in Remark 6 on the reaminder terms related to $\mathcal{E}_j^c$ and $\mathcal{A}_n^c$. □

**Proposition 9.** *The unconditional result in Proposition 8 holds even when $X_0 = X_1$.*

*Proof.* As seen above, we have $\|\langle(\hat{\Gamma}_{h_n}^\dagger - \Gamma_{h_n}^\dagger)X_1, X_1\rangle\| \le C\sum_{j=1}^{h_n} A_j + |\langle r_{2n}X_1, X_1\rangle|\mathbb{I}_{\mathcal{A}_n^c}$, where

$$A_j = \int_{\mathcal{B}_j} |z|^{-1}\|K_n(z)\|_\infty \|G_n(z)\|_\infty \|(zI - \Gamma)^{-1/2}X_1\|^2 dz.$$

This implies that

$$\begin{aligned}
\mathsf{E}[A_j\mathbb{I}_{\mathcal{E}_j}] &\le C\int_{\mathcal{B}_j} \lambda_j^{-1}\mathsf{E}[\|G_n(z)\|_\infty \|(zI-\Gamma)^{-1/2}X_1\|^2]dz \\
&= C\lambda_j^{-1}\sup_{z\in\mathcal{B}_j}\mathsf{E}[\|G_n(z)\|_\infty^2]^{1/2}\mathsf{E}[\|(zI-\Gamma)^{-1/2}X_0\|^4]^{1/2} \\
&\le Cn^{-1/2}(j\log j)^2.
\end{aligned}$$

by Lemmas 1 and 6. We finally have the desired result by the same argument as Proposition 8. □

#### 2.9.2.4 An example of uncorrelated but dependent FPC scores

We provide the proof that $\{\xi_j\}$ constructed in Proposition 1 in the main paper are not independent for the reference.

**Lemma 8.** *Let* $\{W_j\}$ *be a sequence of iid random variables defined as* $\mathsf{P}(W_j = 1) = 1/2 = \mathsf{P}(W_j = -1)$ *and* $\xi \sim \mathsf{N}(0,1)$*, which is independent of* $\{W_j\}$ *and suppose that* $X$ *has the FPC scores in its Karhunen-Loève expansion defined as* $\xi_j = W_j\xi$ *for* $j = 1, 2, \ldots$. *Then, the sequence* $\{\xi_j\}$ *forms a white noise with (uniformly) finite fourth moments, but the random variables in* $\{\xi_j\}$ *are dependent.*

*Proof.* One can show that this example satisfies the condition that the random variables $\xi_j$ are uncorrelated with mean zero, variance one, and finite fourth moments. To see their dependence, assume that $\xi_j$ and $\xi_{j'}$ are independent where $j \neq j'$. Then, by the properties of normal distributions, $\xi_j + \xi_{j'}$ should be normally distributed. However, since $\xi_j + \xi_{j'} = (W_j + W_{j'})\xi$, we have

$$\begin{aligned}
\mathsf{P}(\xi_j + \xi_{j'} = 0) &= \mathsf{P}((W_j + W_{j'} = 0) \cup (\xi = 0)) = 1 - \mathsf{P}((W_j + W_{j'} = 0)^c \cap (\xi = 0)^c) \\
&= 1 - \mathsf{P}((W_j + W_{j'} = 0)^c)\mathsf{P}((\xi = 0)^c) = 1 - \mathsf{P}((W_j + W_{j'} = 0)^c) \\
&= \mathsf{P}(W_j + W_{j'} = 0)) = \mathsf{P}(W_j = 1 = -W_{j'}) + \mathsf{P}(W_j = -1 = -W_{j'}) \\
&= \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{2},
\end{aligned}$$

which is the contradiction to the fact that the normal distribution is continuous. Thus, for each distinct $j, j'$, $\xi_j$ and $\xi_{j'}$ are not independent. □

#### 2.9.2.5 Proof of unconditional CLT

*Proof of the unconditional result on* $X_0$ *of Theorem 1 in the main paper.* From Theorem 1, for each $y \in \mathbb{R}$, we obtain $\mathsf{P}(T_n^{bias}(X_0) \leq y | \mathcal{X}_n, X_0) \xrightarrow{\mathsf{P}} \Phi(y/\sigma_\varepsilon)$ as $n \to \infty$ where

$$T_n^{bias}(X_0) \equiv \sqrt{n/t_{h_n}(X_0)}[\langle \hat{\beta}_{h_n}, X_0 \rangle - \langle \Pi_{h_n}\beta, X_0 \rangle],$$

implying that

$$\mathsf{P}(T_n^{bias}(X_0) \leq y | \mathcal{X}_n) = \mathsf{E}[\mathsf{P}(T_n^{bias}(X_0) \leq y | \mathcal{X}_n, X_0) | \mathcal{X}_n] \xrightarrow{\mathsf{P}} \mathsf{E}[\Phi(y/\sigma_e) | \mathcal{X}_n] = \Phi(y/\sigma_\varepsilon)$$

as $n \to \infty$ due to a subsequence argument (cf. Theorem 20.5 of [4]) and Theorem 9.5.1 of [2]. By using Polya's theorem (Theorem 9.1.4 of [2]) again, as $n \to \infty$, we finally have

$$\sup_{y \in \mathbb{R}} |\mathsf{P}(T_n^{bias}(X_0) \leq y|\mathcal{X}_n) - \Phi(y/\sigma_\varepsilon)| \xrightarrow{\mathsf{P}} 0. \tag{2.21}$$

□

*Proof of Corollary 1 in the main paper.* It can be shown by a similar argument to the above proof for the unconditional CLT on $X_0$. □

#### 2.9.2.6 The unbiased CLT

One can achieve the CLT for unbiased centering $\langle \beta, X_0 \rangle$ as an analog of the residual bootstrap for unbiased centering $\langle \beta, X_0 \rangle$. For this, the non-random bias should be controlled via the smoothness assumption on the slope function $\beta$. Recall that the condition $B(u,v)$ depends on generic constants $v, u > 0$ and a function $m(j,u) \equiv \max\{j^u, \sum_{i=1}^{j} \delta_i^{-2}\}$ of integer $j \geq 1$.
Condition $B(u,v)$ : $\sup_{j \geq 1} \langle \beta, e_j \rangle^2 j^{v-1} m(j,u) < \infty$.

**Lemma 9.** *We see that*

$$\frac{n}{h_n} \mathsf{E}[\langle \Pi_{h_n}\beta - \beta, X_0 \rangle^2] \leq \frac{n}{h_n^v \max\left\{h_n^u, \sum_{j=1}^{h_n} \delta_j^{-2}\right\}} \left( \sum_{j > h_n} \lambda_j \right) \sup_{j \in \mathbb{N}} \left[ \langle \beta, e_j \rangle^2 j^{v-1} m(j,u) \right]$$

*for $v, u > 0$, and hence, under Conditions $B(u,v)$, as $n \to \infty$, if*
$n = O\left(h_n^v \max\left\{h_n^u, \sum_{j=1}^{h_n} \delta_j^{-2}\right\}\right)$, *we have that* $\sqrt{\frac{n}{t_{h_n}(X_0)}} \langle \Pi_{h_n}\beta - \beta, X_0 \rangle \xrightarrow{\mathsf{P}} 0$.

*Proof.* We first notice that $\frac{n}{h_n} \mathsf{E}[\langle \Pi_{h_n}\beta - \beta, X_0 \rangle^2] = \frac{n}{h_n} \sum_{j > h_n} \lambda_j \langle \beta, e_j \rangle^2$. We first see that

$$\begin{aligned} \frac{n}{h_n} \sum_{j > h_n} \lambda_j \langle \beta, e_j \rangle^2 &= \frac{n}{h_n} \sum_{j > h_n} \lambda_j \left( j^{v-1} \sum_{l=1}^{j} \delta_l^{-2} \right)^{-1} \left( j^{v-1} \sum_{l=1}^{j} \delta_l^{-2} \right) \langle \beta, e_j \rangle^2 \\ &\leq \frac{n}{h_n^v \sum_{j=1}^{h_n} \delta_j^{-2}} \left( \sum_{j > h_n} \lambda_j \right) \sup_{j \in \mathbb{N}} \left[ j^{v-1} \left( \sum_{l=1}^{j} \delta_l^{-2} \right) \langle \beta, e_j \rangle^2 \right]. \end{aligned}$$

Also,

$$\frac{n}{h_n}\sum_{j>h_n}\lambda_j\left\langle\beta,e_j\right\rangle^2=\frac{n}{h_n}\sum_{j>h_n}\lambda_j(j^{v-1+u})^{-1}j^{v-1+u}\langle\beta,e_j\rangle^2$$
$$\leq\frac{n}{h_n^{v+u}}\left(\sum_{j>h_n}\lambda_j\right)\sup_{j\in\mathbb{N}}\left[j^{v-1}j^u\langle\beta,e_j\rangle^2\right].$$

We thus have the desired inequality and convergence of the non-random bias term in probability from the assumption $h_n t_{h_n}(X_0)^{-1}=O_{\mathsf{P}}(1)$ and the Chebyshev inequality. □

In what follows, $\tilde{\mathsf{P}}$ denotes either $\mathsf{P}(\cdot|\mathcal{X}_n)$ or $\mathsf{P}(\cdot|\mathcal{X}_n,X_0)$ as Theorem 1 in the main paper.

**Theorem 5** (Unbiased CLT)**.** *Under the assumptions of Theorem 1 in the main paper, suppose that $n=O(m(h_n,u))$ holds with Condition B(u, v) for some $u>5$ and $v>0$. Then, as $n\to\infty$,*

$$\sup_{y\in\mathbb{R}}\left|\tilde{\mathsf{P}}\left(\sqrt{\frac{n}{t_{h_n}(X_0)}}[\langle\hat{\beta}_{h_n},X_0\rangle-\langle\beta,X_0\rangle]\leq y\right)-\Phi(y/\sigma_\varepsilon)\right|\xrightarrow{\mathsf{P}}0,$$

*where the above also holds with the sample version $\hat{t}_{h_n}(X_0)$ of $t_{h_n}(X_0)$.*

*Proof.* It follows from Theorem 1 in the main paper and Lemma 9. □

Under polynomial decay rates on eigengaps $\delta_j\asymp j^{-a}$ (implying $\lambda_j\asymp j^{-a+1}$) and coordinate projections $|\langle\beta,e_j\rangle|\asymp j^{-b}$ for some constants $a>2$ and $b>1$, one can derive the following corollary; here and in the following, we write $r_n\asymp s_n$ if $r_n/s_n$ is bounded away from both zero and infinity for generic sequences $r_n$ and $s_n>0$.

**Corollary 6.** *Under the above polynomial decay rates, suppose (A1)-(A2) along with $h_n t_{h_n}(X_0)^{-1}=O_{\mathsf{P}}(1)$. If $n\asymp h_n^{v_h}$ for some $\max\{5,(2a+1)\}<v_h<a+2b-1$, then the conclusions of Theorem 5 remain valid.*

*Proof.* The dominating term for the non-random bias is

$$\frac{n}{h_n}\sum_{j>h_n}\lambda_j\left\langle\beta,e_j\right\rangle^2\leq Cnh_n^{-1}\sum_{j>h_n}j^{-a-2b}\asymp nh_n^{-a-2b+1}.$$

Suppose that $n \asymp h_n^{v_h}$ where $v_h = \{5 \vee (2a+1)\} + \kappa_h$ for $\kappa_h > 0$. Then, we get the convergences $n^{-1/2} h_n^{5/2} (\log h_n)^2 \to 0$ and $n^{-1} \sum_{j=1}^{h_n} \delta_j^{-2} \asymp n^{-1} h_n^{2a+1} \to 0$ as $n \to \infty$. We thus now check the convergence of the remaining non-random bias term. Since

$$n h_n^{-a-2b+1} = \frac{n}{h_n^{v_h}} h_n^{v_h-(a+2b-1)},$$

if $v_h < a + 2b - 1$, then the upper bound of the non-random bias term satisfies $\frac{n}{h_n} \sum_{j>h_n} \lambda_j \langle \beta, e_j \rangle^2 = o(1)$, and the result follows by Lemma 9. □

### 2.9.3 Validity of the residual bootstrap

To establish the consistency of the bootstrap error distribution, let $F$ be the common cumulative distribution function (CDF) of the errors $\{\varepsilon_i\}_{i=1}^n$. Also, let $F_n$ and $\hat{F}_n$ denote the empirical distributions of the errors $\{\varepsilon_i\}_{i=1}^n$ and the centered residuals $\{\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}\}_{i=1}^n$, respectively.

**Theorem 6.** *As $n \to \infty$, if $\|\hat{\beta}_{k_n} - \beta\| \xrightarrow{\mathsf{P}} 0$, then we have $d_2(\hat{F}, F) \xrightarrow{\mathsf{P}} 0$.*

*Proof.* The proof is along the lines of the proof of Theorem 3.1 in [26]. By Lemma 8.4 of [3], $d_2(F_n, F) \xrightarrow[n\to\infty]{\text{a.s.}} 0$. Note that

$$d_2(F_n, \hat{F}_n)^2 \leq n^{-1} \sum_{i=1}^n \{\varepsilon_i - (\hat{\varepsilon}_i - \bar{\hat{\varepsilon}})\}^2 \leq C \left\{ n^{-1} \sum_{i=1}^n (\varepsilon_i - \hat{\varepsilon}_i)^2 + (\bar{\hat{\varepsilon}})^2 \right\}.$$

Since $(\bar{\hat{\varepsilon}})^2 = (\bar{\hat{\varepsilon}} - \bar{\varepsilon} + \bar{\varepsilon})^2 \leq C\{(\bar{\hat{\varepsilon}} - \bar{\varepsilon})^2 + \bar{\varepsilon}^2\}$ and

$$(\bar{\hat{\varepsilon}} - \bar{\varepsilon})^2 = \left\{ n^{-1} \sum_{i=1}^n (\hat{\varepsilon}_i - \varepsilon_i) \right\}^2 \leq n^{-1} \sum_{i=1}^n (\hat{\varepsilon}_i - \varepsilon_i)^2,$$

we have

$$d_2(F_n, \hat{F}_n)^2 \leq C \left\{ n^{-1} \sum_{i=1}^n (\varepsilon_i - \hat{\varepsilon}_i)^2 + \bar{\varepsilon}^2 \right\}.$$

As $n \to \infty$, since $\bar{\varepsilon} \to \mathsf{E}[\varepsilon_1] = 0$ almost surely, it suffices to show that $n^{-1} \sum_{i=1}^n (\varepsilon_i - \hat{\varepsilon}_i)^2 \xrightarrow{\mathsf{P}} 0$. We observe that

$$n^{-1} \sum_{i=1}^n (\varepsilon_i - \hat{\varepsilon}_i)^2 = n^{-1} \sum_{i=1}^n \langle \hat{\beta}_{k_n} - \beta, X_i \rangle^2 \leq \|\hat{\beta}_{k_n} - \beta\|^2 n^{-1} \sum_{i=1}^n \|X_i\|^2$$

since $\hat{\varepsilon}_i - \varepsilon_i = \langle \hat{\beta}_{k_n} - \beta, X_i \rangle$. As $n \to \infty$, since $n^{-1} \sum_{i=1}^n \|X_i\|^2 \to E[\|X_1\|^2] < \infty$ almost surely, and $\|\hat{\beta}_{k_n} - \beta\| = o_{\mathsf{P}}(1)$, we have the consistency of bootstrap error distribution as follows:

$$\begin{aligned} d_2(\hat{F}_n, F) \le d_2(\hat{F}_n, F_n) + d_2(F_n, F) &\le C\sqrt{n^{-1}\sum_{i=1}^n (\varepsilon_i - \hat{\varepsilon}_i)^2 + \bar{\varepsilon}^2} + d_2(F_n, F) \\ &\le C\sqrt{\|\hat{\beta}_{k_n} - \beta\|^2 \left(n^{-1}\sum_{i=1}^n \|X_i\|^2\right) + \bar{\varepsilon}^2} + d_2(F_n, F) \xrightarrow{\mathsf{P}} 0. \end{aligned}$$

□

The following propisition helps to prove the bootstrap consistency unconditional on $X_0$.

**Proposition 10.** *Proposition 3 holds even when $X_0 = X_1$.*

*Proof.* It follows from Proposition 9 along with the same argument of the proof of Proposition 3 in the main paper. □

*Proof of the unconditional result of Theorem 2 in the main paper.* One can show that as $n \to \infty$,

$$\mathsf{E}^X[\hat{v}_n^2(X_1)] = \sigma_\varepsilon^2 \frac{\hat{t}_{h_n}(X_1)}{t_{h_n}(X_1)} \xrightarrow{\mathsf{P}} \sigma_\varepsilon^2,$$

where $\hat{v}_n(X_1) \equiv \sqrt{n/t_{h_n}(X_1)}\langle \hat{\Gamma}_{h_n}^\dagger U_n, X_1\rangle$, by using Proposition 9 and the same argument in the proof of Theorem 2. Therefore, the argument in the proof of Theorem 2 works even with $X_0 = X_1$ by Propositions 6-7 and 10, and we obtain

$$\sup_{y \in \mathbb{R}} |\mathsf{P}^*(\hat{v}_n^*(X_0^*) \le y | \mathcal{X}_n, X_0^* = X_1) - \Phi(y/\sigma_\varepsilon)| \xrightarrow{\mathsf{P}} 0, \tag{2.22}$$

where

$$\hat{v}_n^*(X_0^*) \equiv \sqrt{n/t_{h_n}(X_0^*)}\langle \hat{\Gamma}_{h_n}^\dagger U_n^*, X_0^*\rangle = \sqrt{n/t_{h_n}(X_0^*)}[\langle \hat{\beta}_{h_n}^*, X_0^*\rangle - \langle \hat{\Pi}_{g_n}\hat{\beta}_{g_n}, X_0^*\rangle].$$

The bootstrap distribution of $\hat{v}_n^*(X_0^*)$ unconditional on $X_0^*$ is given as

$$\hat{G}_n(y) \equiv \mathsf{P}^*(\hat{v}_n^*(X_0^*) \le y|\mathcal{X}_n) = \mathsf{E}^*[\mathsf{P}^*(\hat{v}_n^*(X_0^*) \le y|\mathcal{X}_n, X_0^*)] = n^{-1}\sum_{i=1}^n \hat{G}_n(y|X_i)$$

where $\hat{G}_n(y|X_i) \equiv \mathsf{P}^*(\hat{v}_n^*(X_0^*) \le y|\mathcal{X}_n, X_0^* = X_i)$ denotes the bootstrap distribution of $\hat{v}_n^*(X_0^*)$ conditional on $X_0^* = X_i$. We then have from the bounded convergence theorem and (2.22) that

$$\mathsf{E}\left[\sup_{y\in\mathbb{R}} |\hat{G}_n(y) - \Phi(y/\sigma_\varepsilon)|\right] \le n^{-1}\sum_{i=1}^{n} \mathsf{E}\left[\sup_{y\in\mathbb{R}} |\hat{G}_n(y|X_i) - \Phi(y/\sigma_\varepsilon)|\right]$$
$$= \mathsf{E}\left[\sup_{y\in\mathbb{R}} |\hat{G}_n(y|X_1) - \Phi(y/\sigma_\varepsilon)|\right] \to 0$$

using that $\sup_{y\in\mathbb{R}} |\hat{G}_n(y|X_1) - \Phi(y/\sigma_\varepsilon)| \overset{\mathrm{d}}{=} \sup_{y\in\mathbb{R}} |\hat{G}_n(y|X_i) - \Phi(y/\sigma_\varepsilon)|$ for each $i = 1, \ldots, n$. This implies that

$$\sup_{y\in\mathbb{R}} |\hat{G}_n(y) - \Phi(y/\sigma_\varepsilon)| \overset{\mathsf{P}}{\to} 0. \tag{2.23}$$

Consequently, we have the desired result by (2.21) and (2.23). □

We now only need to show the detailed convergence rates that appear in Theorem 3 in the main paper. The decomposition of the bootstrap bias term is re-stated here:

$$\hat{\Pi}_{h_n}\hat{\beta}_{g_n} - \hat{\Pi}_{g_n} = (\hat{\Pi}_{h_n} - \Pi_{h_n})(\hat{\beta}_{g_n} - \beta) + (\hat{\Pi}_{h_n} - \Pi_{h_n})\beta + (\Pi_{h_n} - I)(\hat{\beta}_{g_n} - \beta) + (\Pi_{h_n} - I)\beta. \tag{2.24}$$

In (2.24), the quantities related to second and fourth terms can be dealt with by Proposition 17 and Lemma 9. In what follows, we investigate the rates of convergence of the quantities related to the first and third terms in (2.24). Before seeing the details, the following two lemmas will be proved.

**Lemma 10.** *Suppose that $g_n > h_n$, $\{t_{g_n}(X_0) - t_{h_n}(X_0)\}^{-1} = O_{\mathsf{P}}(1)$, and $n^{-1/2}(g_n - h_n)^2 \to 0$, as $n \to \infty$. Then, as $n \to \infty$, we have we have*

$$\sup_{y\in\mathbb{R}} \left|\mathsf{P}^X\left(\sqrt{\frac{n}{t_{g_n}(X_0) - t_{h_n}(X_0)}}\left\langle (I - \Pi_{h_n})\Gamma_{g_n}^{\dagger} U_n, X_0\right\rangle \le y\right) - \Phi(y/\sigma_\varepsilon)\right| \overset{\mathsf{P}}{\to} 0.$$

*Proof.* The argument of the proof is similar to that in Proposition 2 in the main paper. Note that $\langle \Gamma_{g_n}^{\dagger} U_n, (I - \Pi_{h_n})X_0\rangle = \sum_{i=1}^{n} Z_{i,n}$ where $Z_{i,n} = n^{-1}\langle \Gamma_{g_n}^{\dagger} X_i, (I - \Pi_{h_n})X_0\rangle\varepsilon_i$. Note that

$\mathsf{E}^X[Z_{i,n}] = 0$ and

$$v_n^2 = \sum_{i=1}^n \mathsf{E}^X[Z_{i,n}^2] = \frac{\sigma_\varepsilon^2}{n^2}\sum_{i=1}^n \left\langle \Gamma_{g_n}^\dagger X_i, (I-\Pi_{h_n})X_0 \right\rangle^2$$
$$= \frac{\sigma_\varepsilon^2}{n}\left\langle \Gamma_n \Gamma_{g_n}^\dagger (I-\Pi_{h_n})X_0, \Gamma_{g_n}^\dagger (I-\Pi_{h_n})X_0 \right\rangle$$
$$= \frac{\sigma_\varepsilon^2}{n}(A_n + t_{g_n}((I-\Pi_{h_n})X_0))$$

where $A_n \equiv \langle (\Gamma_n - \Gamma)\Gamma_{g_n}^\dagger (I-\Pi_{h_n})X_0, \Gamma_{g_n}^\dagger (I-\Pi_{h_n})X_0\rangle$ and

$$t_{g_n}((I-\Pi_{h_n})X_0) = \left\langle \Gamma_{g_n}^\dagger (I-\Pi_{h_n})X_0, (I-\Pi_{h_n})X_0 \right\rangle.$$

with

$$\Gamma_{g_n}^\dagger (I-\Pi_{h_n})X_0 = \sum_{j=1}^{g_n} \lambda_j^{-1}\langle (I-\Pi_{h_n})X_0, e_j\rangle e_j = \sum_{j=1}\sum_{l>h_n} \lambda_j^{-1}\langle X_0, e_l\rangle \langle e_l, e_j\rangle e_j$$
$$= \sum_{j>h_n}^{g_n} \lambda_j^{-1}\langle X_0, e_j\rangle e_j.$$

We can write $A_n = n^{-1}\sum_{i=1}^n B_{i,n}$ where

$$B_{i,n} \equiv \left\langle (X_i \otimes X_i - \Gamma)\Gamma_{g_n}^\dagger (I-\Pi_{h_n})X_0, \Gamma_{g_n}^\dagger (I-\Pi_{h_n})X_0 \right\rangle.$$

By applying the same argument as the proof of Lemma 7, we have

$B_{i,n} = \sum_{j,l>h_n}^{g_n} \xi_{0j}\xi_{0l}\{\xi_{ij}\xi_{il} - \mathbb{I}(j=l)\}$, and thus,

$$E[A_n^2] \le Cn^{-1}(g_n - h_n)^4,$$

which implies that $A_n = O_P(n^{-1/2}(g_n-h_n)^2)$. We next observe that

$$t_{g_n}((I-\Pi_{h_n})X_0) = \left\langle \sum_{j>h_n}^{g_n} \lambda_j^{-1}\langle X_0, e_j\rangle e_j, \sum_{l>h_n}\langle X_0, e_l\rangle e_l \right\rangle = \sum_{j>h_n}^{g_n} \lambda_j^{-1}\langle X_0, e_j\rangle^2$$
$$t_{g_n}(X_0) - t_{h_n}(X_0).$$

Then, $E[t_{g_n}((I-\Pi_{h_n})X_0)] = g_n - h_n$, which implies that $t_{g_n}((I-\Pi_{h_n})X_0) = O_P(g_n - h_n)$. As $n \to \infty$, since $\{t_{g_n}(X_0) - t_{h_n}(X_0)\}^{-1} = O_{\mathsf{P}}(1)$ and $n^{-1/2}(g_n-h_n)^2 \to 0$ by the assumptions, we have

$$\frac{t_{g_n}((I-\Pi_{h_n})X_0)}{A_n + t_{g_n}((I-\Pi_{h_n})X_0)} = 1 + o_P(1),$$

and the Lindeberg condition is satisfied by applying the same argument as the proof of Proposition 2 in the main paper. Since

$$v_n^{-1}\sum_{i=1}^{n} Z_{i,n} = \sqrt{\frac{n}{\sigma_\varepsilon(A_n + t_{g_n}((I-\Pi_{h_n})X_0))}}\left\langle \Gamma_{g_n}^{\dagger} U_n, (I-\Pi_{h_n})X_0\right\rangle$$
$$= \sqrt{\frac{t_{g_n}((I-\Pi_{h_n})X_0)}{A_n + t_{g_n}((I-\Pi_{h_n})X_0)}}\sqrt{\frac{n}{\sigma_\varepsilon t_{g_n}((I-\Pi_{h_n})X_0)}}\left\langle \Gamma_{g_n}^{\dagger} U_n, (I-\Pi_{h_n})X_0\right\rangle,$$

we finally have

$$\sup_{y\in\mathbb{R}}\left|\mathsf{P}^X\left(\sqrt{\frac{n}{t_{g_n}(X_0)-t_{h_n}(X_0)}}\left\langle (I-\Pi_{h_n})\Gamma_{g_n}^{\dagger} U_n, X_0\right\rangle \le y\right) - \Phi(y/\sigma_\varepsilon)\right| \xrightarrow{\mathsf{P}} 0$$

as $n\to\infty$. □

The second term in (2.24) is bounded as follows.

**Proposition 11.** *Suppose that $g_n > h_n$ with $h_n\to\infty$ as $n\to\infty$. Then, as $n\to\infty$, we have*

$$\mathsf{E}^X\left[\sqrt{\frac{n}{t_{h_n}(X_0)}}|\langle(\hat{\Pi}_{h_n}-\Pi_{h_n})(\hat{\beta}_{g_n}-\beta), X_0\rangle|\right]$$
$$=O_{\mathsf{P}}\left(M_{n,g_n}h_n^{-1/2}\sum_{j=1}^{h_n} j\log j\right) + O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sqrt{\sum_{j>g_n}\beta_j^2\sum_{j=1}^{h_n}(j\log j)^2}\right) + o_{\mathsf{P}}(1),$$

*where for integer $j\ge 1$, $M_{n,j}$ is defined as*

$$M_{n,j} = n^{-1}\sum_{l=1}^{j}\delta_l^{-1/2}(l\log l)^{3/2} + n^{-1/2}\left(\sum_{l=1}^{j}\lambda_l^{-1}\right)^{1/2} + n^{-1/2}\sum_{l=1}^{j} l\log l. \tag{2.25}$$

*Therefore, if $n^{-1}g_n^4(\log g_n)^2h_n^3(\log h_n)^2\to 0$ (which is implied by $g_n^{v_g}=O(n)$ for some $v_g>7$), then for each $\eta>0$,*

$$\mathsf{P}^X\left(\sqrt{\frac{n}{t_{h_n}(X_0)}}|\langle(\hat{\Pi}_{h_n}-\Pi_{h_n})(\hat{\beta}_{g_n}-\beta), X_0\rangle| > \eta\right)\xrightarrow{\mathsf{P}} 0.$$

*Proof.* Following the spirit of Remark 6, we ignore the remainder terms related to either $\mathcal{E}_j^c$ or $\mathcal{A}_n^c$. Based on Lemmas 3-5 and the decomposition (3.38), one can see that

$$\begin{aligned}
&\mathsf{E}^X\left[\sqrt{\frac{n}{t_{h_n}(X_0)}}|\langle(\hat{\Pi}_{h_n}-\Pi_{h_n})(\hat{\beta}_{g_n}-\Pi_{h_n}\beta),X_0\rangle|\right]\\
=&\mathsf{E}^X\left[\sqrt{\frac{n}{t_{h_n}(X_0)}}|\langle(\hat{\beta}_{g_n}-\Pi_{h_n}\beta),(\hat{\Pi}_{h_n}-\Pi_{h_n})X_0\rangle|\right]\\
\leq&\mathsf{E}^X\left[\sqrt{\frac{n}{t_{h_n}(X_0)}}\|(\hat{\beta}_{g_n}-\Pi_{h_n}\beta)\|\|(\hat{\Pi}_{h_n}-\Pi_{h_n})X_0\|\right]\\
=&O_{\mathsf{P}}\left(M_{n,g_n}h_n^{-1/2}\sum_{j=1}^{h_n}j\log j\right).
\end{aligned}$$

Meanwhile, to bound the remaining part related to $\langle(\hat{\Pi}_{h_n}-\Pi_{h_n})(I-\Pi_{g_n}),X_0\rangle$, we have

$$\begin{aligned}
\hat{\Pi}_{h_n}-\Pi_{h_n}&=\frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}\left\{(zI-\Gamma_n)^{-1}-(zI-\Gamma)^{-1}\right\}dz\\
&=\mathcal{S}_n+\mathcal{R}_n+r_{1n}\mathbb{I}_{\mathcal{A}_n^c}
\end{aligned}$$

as seen in the proof of Proposition 2 in [CMS], where

$$\begin{aligned}
\mathcal{S}_n&=\frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}(zI-\Gamma)^{-1}(\Gamma_n-\Gamma)(zI-\Gamma)^{-1}dz,\\
\mathcal{R}_n&=\frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}(zI-\Gamma)^{-1}(\Gamma_n-\Gamma)(zI-\Gamma)^{-1}(\Gamma_n-\Gamma)(zI-\Gamma_n)^{-1}dz.
\end{aligned}$$

Following the proof of Proposition 2 in [CMS], as $n\to\infty$, one can show that

$$\begin{aligned}
\frac{n}{h_n}E\left[\langle\mathcal{S}_n(I-\Pi_{g_n})\beta,X_0\rangle^2\right]&\leq Ch_n^{-1}\sum_{l=1}^{h_n}\lambda_l\left(\sum_{l'>g_n}|\beta_{l'}|\frac{\lambda_l^{1/2}\lambda_{l'}^{1/2}}{\lambda_l-\lambda_{l'}}\right)^2\\
&\leq Ch_n^{-1}\sum_{l=1}^{h_n}\lambda_l\left(\sum_{l'>h_n}|\beta_{l'}|\frac{\lambda_l^{1/2}\lambda_{l'}^{1/2}}{\lambda_l-\lambda_{l'}}\right)^2\to 0,
\end{aligned}$$

which implies that $\sqrt{\frac{n}{h_n}}\langle\mathcal{S}_n(I-\Pi_{g_n})\beta,X_0\rangle\xrightarrow{\mathsf{P}}0$. Second, note that $\|(I-\Pi_{g_n})\beta\|\leq\sqrt{\sum_{j>g_n}\beta_j^2}$. Thus, following the proof of Proposition 2 in [CMS], as $n\to\infty$, we have

$$\sqrt{\frac{n}{h_n}}\langle\mathcal{R}_n(I-\Pi_{g_n})\beta,X_0\rangle=O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sqrt{\sum_{j>g_n}\beta_j^2}\sum_{j=1}^{h_n}(j\log j)^2\right)+o_{\mathsf{P}}(1),$$

where the latter $o_{\mathsf{P}}(1)$ is related to the event sets $\mathcal{E}_j^c$. Hence, as $n \to \infty$, we have that

$$\sqrt{\frac{n}{t_{h_n}(X_0)}}\left|\langle(\hat{\Pi}_{h_n} - \Pi_{h_n})(I - \Pi_{g_n})\beta, X_0\rangle\right| = o_{\mathsf{P}}(1) + O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sqrt{\sum_{j>g_n}\beta_j^2}\sum_{j=1}^{h_n}(j\log j)^2\right).$$

It now suffice to show that the convergence rete $n^{-1}g_n^4(\log g_n)^2h_n^3(\log h_n)^2 \to 0$ implies the result. We observe from Cauchy-Schwarz inequality that as $n \to \infty$, if $n^{-1}\sum_{j=1}^{g_n}\delta_j^{-2} \to 0$, then

$$\left(n^{-1/2}g_n^{-1/2}\sum_{j=1}^{g_n}\delta_j^{-1}\right)^2 = n^{-1}g_n^{-1}\left(\sum_{j=1}^{g_n}\delta_j^{-1}\right)^2 \leq n^{-1}\sum_{j=1}^{g_n}\delta_j^{-2} \to 0,$$
$$\left(n^{-1/4}g_n^{-3/4}\sum_{j=1}^{g_n}\delta_j^{-1/2}\right)^2 = n^{-1/2}g_n^{-3/2}\left(\sum_{j=1}^{g_n}\delta_j^{-1/2}\right)^2 \leq n^{-1/2}g_n^{-1/2}\sum_{j=1}^{g_n}\delta_j^{-1} \to 0.$$

From these, we have the desired result by showing that each term in $M_{n,g_n}h_n^{-1/2}\sum_{j=1}^{h_n}\log h_n$ is dominated by $n^{-1}g_n^4(\log g_n)^2h_n^3(\log h_n)^2$ as follows.

1. The first term in $M_{n,g_n}h_n^{-1/2}\sum_{j=1}^{h_n}\log h_n$ is bounded by

$$\begin{aligned}
&n^{-1}\left\{\sum_{j=1}^{g_n}\delta_j^{-1/2}(j\log j)^{3/2}\right\}\left\{h_n^{-1/2}\sum_{j=1}^{h_n}j\log j\right\}\\
\leq &n^{-1}g_n^{3/2}(\log g_n)^{3/2}\left(\sum_{j=1}^{g_n}\delta_j^{-1/2}\right)h_n^{3/2}\log h_n\\
= &\left(n^{-1/4}g_n^{-3/4}\sum_{j=1}^{g_n}\delta_j^{-1/2}\right)\left\{n^{-3/4}g_n^{9/4}(\log g_n)^{3/2}h_n^{3/2}\log h_n\right\}\\
= &o(1)\left\{\frac{g_n^3(\log g_n)^2h_n^2(\log h_n)^{4/3}}{n}\right\}^{3/4}.
\end{aligned}$$

2. The second term in $M_{n,g_n} h_n^{-1/2} \sum_{j=1}^{h_n} \log h_n$ is bounded by

$$
\begin{aligned}
& n^{-1/2} \left( \sum_{j=1}^{g_n} \lambda_j^{-1} \right)^{1/2} \left\{ h_n^{-1/2} \sum_{j=1}^{h_n} j \log j \right\} \\
\leq & n^{-1/2} \left( \sum_{j=1}^{g_n} \delta_j^{-1} \right)^{1/2} h_n^{3/2} \log h_n \\
= & \left( n^{-1/2} g_n^{-1/2} \sum_{j=1}^{g_n} \delta_j^{-1} \right)^{1/2} n^{-1/4} g_n^{1/4} h_n^{3/2} \log h_n \\
= & o(1) \left\{ \frac{g_n h_n^6 (\log h_n)^4}{n} \right\}^{1/4}.
\end{aligned}
$$

3. The third term in $M_{n,g_n} h_n^{-1/2} \sum_{j=1}^{h_n} \log h_n$ is bounded by

$$
\begin{aligned}
& n^{-1/2} \left\{ \sum_{j=1}^{g_n} j \log j \right\} \left\{ h_n^{-1/2} \sum_{j=1}^{h_n} j \log j \right\} \\
\leq & n^{-1/2} g_n^2 \log g_n h_n^{3/2} \log h_n \\
= & \left\{ \frac{g_n^4 (\log g_n)^2 h_n^3 (\log h_n)^2}{n} \right\}^{1/2}.
\end{aligned}
$$

□

The third term in (2.24) is bounded as follows.

**Proposition 12.** *Suppose that $g_n > h_n$ with $h_n \to \infty$ as $n \to \infty$. As $n \to \infty$, if $g_n/h_n \to 1$, we have that*

$$
\begin{aligned}
& \mathsf{E}^X \left[ \sqrt{\frac{n}{t_{h_n}(X_0)}} |\langle (I - \Pi_{h_n})(\hat{\beta}_{g_n} - \beta), X_0 \rangle| \right] \\
= & O_{\mathsf{P}} \left( n^{-1/2} h_n^{-1/2} \sum_{j=1}^{g_n} (j \log j)^2 \right) + o_{\mathsf{P}}(1) + O_{\mathsf{P}} \left( \sqrt{\frac{n}{h_n} \sum_{j > g_n} \lambda_j \beta_j^2} \right).
\end{aligned}
$$

*Therefore, if further, $n^{-1/2} h_n^{5/2} (\log h_n)^2 \to 0$ (implied by $n^{-1/2} g_n^{5/2} (\log g_n)^2 \to 0$) and Condition $B(u,v)$ are satisfied, then for each $\eta > 0$,*

$$
\mathsf{P}^X \left( \sqrt{\frac{n}{t_{h_n}(X_0)}} |\langle (I - \Pi_{h_n})(\hat{\beta}_{g_n} - \beta), X_0 \rangle| > \eta \right) \xrightarrow{\mathsf{P}} 0.
$$

*Proof.* One can imagine that this can be verifed in a similar manner to the proof of the generalized CLT because the target quantity is here the projection onto a truncated new predictor $(I-\Pi_{h_n})X_0$ at the truncation level $h_n$. We again study each term in this quantity based on the decomposition (3.38).

1. The term $\sqrt{n/t_{h_n}(X_0)}\langle(I-\Pi_{h_n})(\hat{\Gamma}^{\dagger}_{g_n}-\Gamma^{\dagger}_{g_n})U_n, X_0$ is bounded as follows. Note that $(I-\Pi_{h_n})X_0=\sum_{l>h_n}\langle X_0,e_l\rangle e_l$. One can see that

$$\begin{aligned}
&\|(zI-\Gamma)^{-1/2}(I-\Pi_{h_n})X_0\|^2\\
=&\left\|\sum_{l>h_n}\langle X_0,e_l\rangle(zI-\Gamma)^{-1/2}e_l\right\|^2=\left\|\sum_{l>h_n}\langle X_0,e_l\rangle(z-\lambda_l)^{-1/2}e_l\right\|^2\\
=&\sum_{l>h_n}\langle X_0,e_l\rangle^2(z-\lambda_l)^{-1}=\sum_{l>h_n}\frac{\lambda_l\xi_l^2}{|z-\lambda_l|},
\end{aligned}$$

which implies that

$$\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}(I-\Pi_{h_n})X_0\|^2\right]\le Cj\log j$$

as done by Lemma 1. Then, by a similar argument to the proof of Proposition 18, we have

$$\mathsf{E}^X\left[\sqrt{\frac{n}{t_{h_n}(X_0)}}\langle(I-\Pi_{h_n})(\hat{\Gamma}^{\dagger}_{g_n}-\Gamma^{\dagger}_{g_n})U_n,X_0\rangle\right]=O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{g_n}(j\log j)^2\right),$$

$$\sqrt{\frac{n}{t_{h_n}(X_0)}}\langle(I-\Pi_{h_n})(\hat{\Gamma}^{\dagger}_{g_n}-\Gamma^{\dagger}_{g_n})U_n,X_0\rangle=O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{g_n}(j\log j)^2\right).$$

2. The term $\sqrt{n/t_{h_n}(X_0)}\langle(I-\Pi_{h_n})\Gamma^{\dagger}_{g_n}U_n,X_0\rangle$ is bounded as follows. Notice that

$$\begin{aligned}
&\sqrt{\frac{n}{t_{h_n}(X_0)}}\langle(I-\Pi_{h_n})\Gamma^{\dagger}_{g_n}U_n,X_0\rangle\\
=&\sqrt{\frac{t_{g_n}(X_0)-t_{h_n}(X_0)}{t_{h_n}(X_0)}}\sqrt{\frac{n}{t_{g_n}(X_0)-t_{h_n}(X_0)}}\langle\Gamma^{\dagger}_{g_n}U_n,(I-\Pi_{h_n})X_0\rangle\\
=&\sqrt{\frac{t_{g_n}(X_0)}{t_{h_n}(X_0)}-1}\sqrt{\frac{n}{t_{g_n}(X_0)-t_{h_n}(X_0)}}\langle\Gamma^{\dagger}_{g_n}U_n,(I-\Pi_{h_n})X_0\rangle.
\end{aligned}$$

Suppose that $g_n/h_n\to 1$ as $n\to\infty$. Then,

$$\frac{t_{g_n}(X_0)}{t_{h_n}(X_0)}-1=\frac{h_n}{t_{h_n}(X_0)}\frac{t_{g_n}(X_0)-t_{h_n}(X_0)}{g_n-h_n}\frac{g_n-h_n}{h_n}=O_{\mathsf{P}}(1)O_{\mathsf{P}}(1)o(1)=o_{\mathsf{P}}(1).$$

This implies from Lemma 10 that for any $\eta > 0$,

$$\mathsf{P}^X \left( \sqrt{\frac{n}{t_{h_n}(X_0)}} \left| \langle (I - \Pi_{h_n}) \Gamma_{g_n}^\dagger U_n, X_0 \rangle \right| > \eta \right) \xrightarrow[n\to\infty]{} 0$$

as $n \to \infty$.

3. The term $\sqrt{n/t_{h_n}(X_0)} \langle (I - \Pi_{h_n})(\hat{\Pi}_{g_n} - \Pi_{g_n})\beta, X_0 \rangle$ is bounded as follows. As seen in the proof of Proposition 17, we have

$$\begin{aligned} \hat{\Pi}_{g_n} - \Pi_{g_n} &= \frac{1}{2\pi\iota} \sum_{j=1}^{g_n} \int_{\mathcal{B}_j} \left\{ (zI - \Gamma_n)^{-1} - (zI - \Gamma)^{-1} \right\} dz \\ &= \mathcal{S}_{g_n} + \mathcal{R}_{g_n} + r_{1n} \mathbb{I}_{\mathcal{A}_n^c}, \end{aligned}$$

where

$$\begin{aligned} \mathcal{S}_{g_n} &= \frac{1}{2\pi\iota} \sum_{j=1}^{g_n} \int_{\mathcal{B}_j} (zI - \Gamma)^{-1} (\Gamma_n - \Gamma)(zI - \Gamma)^{-1} dz, \\ \mathcal{R}_{g_n} &= \frac{1}{2\pi\iota} \sum_{j=1}^{g_n} \int_{\mathcal{B}_j} (zI - \Gamma)^{-1} (\Gamma_n - \Gamma)(zI - \Gamma)^{-1} (\Gamma_n - \Gamma)(zI - \Gamma_n)^{-1} dz. \end{aligned}$$

Following the proof of Proposition 2 in [CMS], we have

$$\begin{aligned} &\frac{n}{h_n} \mathsf{E} \left[ \langle \mathcal{S}_{g_n} \beta, (I - \Pi_{h_n}) X_0 \rangle^2 \right] \\ \leq & C h_n^{-1} \sum_{l > h_n}^{g_n} \lambda_l \left( \sum_{l' > g_n} |\beta_{l'}| \sqrt{\frac{\lambda_l \lambda_{l'}}{\lambda_l - \lambda_{l'}}} \right)^2 + C h_n^{-1} \sum_{l > g_n} \lambda_l \left( \sum_{l' \leq g_n} |\beta_{l'}| \frac{\sqrt{\lambda_l \lambda_{l'}}}{\lambda_l - \lambda_{l'}} \right)^2. \end{aligned}$$

If $g_n/h_n = O(1)$, then as $n \to \infty$, we have that $\frac{n}{h_n} \mathsf{E} \left[ \langle \mathcal{S}_{g_n} \beta, (I - \Pi_{h_n}) X_0 \rangle^2 \right] \to 0$ as seen in their proof, which implies that $\sqrt{\frac{n}{h_n}} \langle \mathcal{S}_{g_n} \beta, (I - \Pi_{h_n}) X_0 \xrightarrow{\mathsf{P}} 0$. Next, note that $(I - \Pi_{h_n}) X_0 = \sum_{l > h_n} \langle X_0, e_l \rangle e_l$. One can see that

$$\begin{aligned} &\| (zI - \Gamma)^{-1/2} (I - \Pi_{h_n}) X_1 \|^2 \\ = &\left\| \sum_{l > h_n} \langle X_1, e_l \rangle (zI - \Gamma)^{-1/2} e_l \right\|^2 = \left\| \sum_{l > h_n} \langle X_1, e_l \rangle (z - \lambda_l)^{-1/2} e_l \right\|^2 \\ = &\sum_{l > h_n} \langle X_1, e_l \rangle^2 (z - \lambda_l)^{-1} = \sum_{l > h_n} \frac{\lambda_l \xi_l^2}{|z - \lambda_l|} \\ \leq &\sum_{l=1}^{\infty} \frac{\lambda_l \xi_l^2}{|z - \lambda_l|} \end{aligned}$$

which implies that

$$\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}(I-\Pi_{h_n})X_1\|^2\right]\le Cj\log j$$

in a similar fashion to Lemma 1. Thus, following the proof of Proposition 2 in [CMS], as $n\to\infty$, we have

$$\sqrt{\frac{n}{h_n}}\langle\mathcal{R}_{g_n}\beta,(I-\Pi_{h_n})X_0\rangle=O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{g_n}(j\log j)^2\right)+o_{\mathsf{P}}(1).$$

By (a) and (b), if $g_n/h_n=O(1)$, we have

$$\sqrt{\frac{n}{t_{h_n}(X_0)}}\langle(I-\Pi_{h_n})(\hat{\Pi}_{g_n}-\Pi_{g_n})\beta,X_0\rangle=o_{\mathsf{P}}(1)+O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{g_n}(j\log j)^2\right).$$

4. The term $\sqrt{n/t_{h_n}(X_0)}\langle(I-\Pi_{h_n})(\Pi_{g_n}-I)\beta=-(I-\Pi_{h_n})(I-\Pi_{g_n})\beta,X_0\rangle$ is bounded as follows. Notice that $(I-\Pi_{h_n})(I-\Pi_{g_n})=I-\Pi_{h_n\vee g_n}=I-\Pi_{g_n}$ since $g_n>h_n$. Since $\langle(I-\Pi_{g_n})\beta,X_0\rangle=\sum_{j>g_n}\beta_j\langle X_0,e_j\rangle$, we have

$$\mathsf{E}\left[|\langle(I-\Pi_{h_n})(\Pi_{g_n}-I)\beta,X_0\rangle|\right]\le\sqrt{\mathsf{E}\left[|\langle(I-\Pi_{g_n})\beta,X_0\rangle|^2\right]}=\sqrt{\sum_{j>g_n}\lambda_j\beta_j^2},$$

which gives us

$$\sqrt{\frac{n}{t_{h_n}(X_0)}}\langle(I-\Pi_{h_n})(\Pi_{g_n}-I)\beta,X_0\rangle=O_{\mathsf{P}}\left(\sqrt{\frac{n}{h_n}\sum_{j>g_n}\lambda_j\beta_j^2}\right).$$

□

*Proof of Corollary 2 in the main paper.* We first notice that $k_n^{v_k}=O(n)$ implies that $n^{-1/2}k_n^2(\log k_n)\to 0$ and $n^{-1}\sum_{j=1}^{k_n}\delta_j^{-2}\asymp n^{-1}k_n^{2a+1}\to 0$ as $n\to\infty$. This guarantees the consistency of bootstrap error distribution in Theorem 6. The first part under $g_n\le h_n$ follows from the same argument as the proof of Corollary 6. We apply the same argument for the second part. The dominating term for the non-random bias is here

$$\frac{n}{h_n}\sum_{j>g_n}\lambda_j\langle\beta,e_j\rangle^2\le Cng_n^{-1}\sum_{j>g_n}j^{-a-2b}\asymp ng_n^{-a-2b+1}.$$

Suppose that $n \asymp g_n^{v_g}$ where $v_g = \{7 \vee (2a+1)\} + \kappa_g$ for $\kappa_g > 0$. Then, we get the convergences $n^{-1/2} g_n^{7/2} (\log g_n)(\log h_n) \to 0$ and $n^{-1} \sum_{j=1}^{g_n} \delta_j^{-2} \asymp n^{-1} g_n^{2a+1} \to 0$ as $n \to \infty$. Since

$$n g_n^{-a-2b+1} = \frac{n}{g_n^{v_g}} g_n^{v_g-(a+2b-1)},$$

if $v_g < a+2b-1$, then it holds that $\frac{n}{h_n} \sum_{j>g_n} \lambda_j \langle \beta, e_j \rangle^2 = o(1)$, and the result follows by Lemma 9. □

*Proofs of Corollary 3 in the main paper.* One can show the asymptotic normality and the bootstrap consistency for any linear combination $\sum_{i_0=1}^{n_0} a_{i_0} X_{i_0}$ following the same argument of those for single new predictor. The result of Corollary 3 then follows by the Wold device (Theorem 10.4.5 of [2]) and the continuous mapping theorem (Theorem 9.4.2 of [2]). □

*Proof of Corollary 4 in the main paper.* We first consider the individaul prediction of response $Y_0$ with single new predictor $X_0$. To establish the result, the proof is based on a subsequence argument (cf. Theorem 20.5 of [4]). Let $\{n'\} \subseteq \{n\}$ be a subsequence of $\{n\}$. Then, due to Theorems 5-6 and Theorem 3 in the main paper (where distributional convergence holds conditionally on $X_0$ and $X_0^* = X_0$ and thereby also when removing conditioning on $X_0$), there exists a further subsequence $\{n''\}$ and an almost sure event $D \in \mathcal{F}$ such that $\sup_{y\in\mathbb{R}} |\mathsf{P}(T_{n''} \leq y | \mathcal{X}_{n''}) - \Phi(y/\sigma_\varepsilon)| \to 0$, $\sup_{y\in\mathbb{R}} |\mathsf{P}^*(T_{n''}^* \leq y | \mathcal{X}_{n''}) - \Phi(y/\sigma_\varepsilon)| \to 0$, $\sqrt{n''/t_{h_{n''}}(X_0)} \to \infty$, and $d_2(\hat{F}_{n''}, F) \to 0$ as $n'' \to \infty$ on $D$, where

$$T_{n''} \equiv \sqrt{n/t_{h_{n''}}(X_0)}[\langle \hat{\beta}_{h_{n''}}, X_0 \rangle - \langle \beta, X_0 \rangle],$$
$$T_{n''}^* \equiv \sqrt{n/t_{h_{n''}}(X_0)}[\langle \hat{\beta}_{h_{n''}}^*, X_0 \rangle - \langle \hat{\beta}_{g_{n''}}, X_0 \rangle],$$

and $\hat{F}_{n''}$ denotes the empirical distribution function of the centered residuals, while $F$ denotes the distribution function of an error $\varepsilon_0$. It then holds that $\langle \hat{\beta}_{h_{n''}}, X_0 \rangle - \langle \beta, X_0 \rangle \xrightarrow{\mathsf{d}} 0$ and $\langle \hat{\beta}_{h_{n''}}^*, X_0^* \rangle - \langle \hat{\beta}_{g_{n''}}, X_0^* \rangle \xrightarrow{\mathsf{d}} 0$ along the sequences $\{\mathsf{P}(\cdot|\mathcal{X}_{n''})\}$ and $\{\mathsf{P}^*(\cdot|\mathcal{X}_{n''})\}$, respectively, on $D$. This implies by Slutksy's theorem (Theorem 9.1.6 of [2]) that

$$Y_0 - \hat{Y}_0 = \langle \beta, X_0 \rangle - \langle \hat{\beta}_{h_{n''}}, X_0 \rangle + \varepsilon_0 \xrightarrow{\mathsf{d}} \varepsilon_0,$$
$$Y_0^* - \hat{Y}_0^* = \langle \hat{\beta}_{g_{n''}}, X_0 \rangle - \langle \hat{\beta}_{h_{n''}}^*, X_0 \rangle + \varepsilon_0^* \xrightarrow{\mathsf{d}} \varepsilon_0,$$

as $n'' \to \infty$ along the sequences $\{\mathsf{P}(\cdot|\mathcal{X}_{n''})\}$ and $\{\mathsf{P}^*(\cdot|\mathcal{X}_{n''})\}$, respectively, on $D$, using that the conditional distribution of $\varepsilon_0$ given $\mathcal{X}_n$ is $F$ (the unconditional distribution of $\varepsilon_0$) by independence while the bootstrap distribution of $\varepsilon_0^*$ is $\hat{F}_{n''}$ (which converges weakly to $F$). Since convergence in the Levy metric is equivalent to weak convergence (cf. Problem 9.18 of [2]), we have

$$d_L(Y_0 - \hat{Y}_0|\mathcal{X}_{n''}, Y_0^* - \hat{Y}_0^*|\mathcal{X}_{n''}) \leq d_L(Y_0 - \hat{Y}_0|\mathcal{X}_{n''}, \varepsilon_0) + d_L(Y_0^* - \hat{Y}_0^*|\mathcal{X}_{n''}, \varepsilon_0) \to 0$$

as $n'' \to 0$ on $D$. Due to Theorem 20.5 of [4], we derive $d_L(Y_0 - \hat{Y}_0|\mathcal{X}_n, Y_0^* - \hat{Y}_0^*|\mathcal{X}_n) \xrightarrow{\mathsf{P}} 0$ as $n \to \infty$. Simultaneous prediction result follows from the same argument as the proof of Corollary 3. □

## 2.10 Additional simulation results

### 2.10.1 Other scenarios under consideration

In addition to the set-ups for simulation described in Section 2.5.1 of the main paper, we consider the following choices of FPC scores to obtain uncorrelated (but possibly dependent) sequence and error distributions:

(FPC1) Define $\xi_j = \xi W_j$ where $\xi \sim \mathsf{N}(0,1)$ and $W_j$ are iid with $\mathsf{P}(W_j = 1) = 1/2 = \mathsf{P}(W_j = -1)$.

(FPC2) Define $\xi_j = \xi W_j$ where $\xi \sim \mathsf{N}(0,1)$, $W_j \overset{\text{iid}}{\sim} \mathsf{N}(0,1)$, and $\xi$ and $\{W_j\}$ are independent.

(FPC3) let $\xi_j = V_j W_j$, where $\{W_j\}$ are iid $\mathsf{N}(0,1)$ variables and, independently, let $\{V_j\}$ be a stationary autoregressive process such that each $V_j \sim \mathsf{N}(0,1)$ and $V_{j+1}|V_j \sim \mathsf{N}(0.5V_j, 1.5)$.

(E1) $\varepsilon_i \overset{\text{iid}}{\sim} \mathsf{N}(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon^2 = 2$.

(E2) $\varepsilon_i \overset{\text{iid}}{\sim} t(\nu_\varepsilon)$ with $\sigma_\varepsilon^2 = 2$ and $\nu_\varepsilon = 4$.

(E3) $\varepsilon_i \overset{\text{iid}}{\sim} U(-a, a)$ with $a = \sqrt{6}$ so that $\sigma_\varepsilon^2 = 2$.

It can be shown that the random variables in the sequence $\{\xi_j\}$ are uncorrelated but not independent because $\mathsf{E}[\xi_j^2 \xi_{j'}^2] \neq \mathsf{E}[\xi_j^2]\mathsf{E}[\xi_{j'}^2]$ for each distinct $j, j' \in \mathbb{N}$ Since all the results show a similar pattern, we report only the results from the third types (FPC3) and (E3) of FPC scores and error term, respectively, considered above.

### 2.10.2 Sampling distributions

The finite sampling distributions of the target quantities $\sqrt{n/t_{h_n}(X_0)}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ and $\sqrt{n/h_n}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ are investigated in this section based on the scenarios described as above and in Section 2.5.1 of the main paper. In addition to Figure 2.1 of the main paper, Figures-2.8-2.11 show the kernel-estimated densities of these sampling distributions with different choice of tuning parameters $h_n \in \{1, \ldots, 15\}$ and $M = 1000$ Monte Carlo simulation size.

For each $m = 1, \ldots, M$, perform the following.

1. Simulate independent $X_1, \ldots, X_n$ with $X_i \overset{\mathrm{d}}{=} X$ and $\varepsilon_i \overset{\mathrm{iid}}{\sim} N(0,1)$ for $i = 1, \ldots, n$, and independently simulate $X_0 \overset{\mathrm{d}}{=} X$ and $\varepsilon_0 \sim N(0,1)$. Then, generate $Y_1, \ldots, Y_n$ and $Y_0$ as $Y_i = \langle\beta, X_i\rangle + \varepsilon_i$ for $i = 1, \ldots, n$ and $Y_0 = \langle\beta, X_0\rangle = \varepsilon_0$.

2. Compute $\Gamma^{\dagger}_{h_n}$, $\Pi^{\dagger}_{h_n}$, $\hat{\Gamma}^{\dagger}_{h_n}$, $\hat{\Pi}_{h_n}$, $\Delta_n$, $U_n$, and $\hat{\beta}_{h_n}$ to get the components in the decomposition of $\hat{\beta}_{h_n} - \beta$:

$$\hat{\beta}_{h_n} - \beta = (\hat{\Gamma}^{\dagger}_{h_n} - \Gamma^{\dagger}_{h_n})U_n + \Gamma^{\dagger}_{h_n}U_n + (\hat{\Pi}_{h_n} - \Pi_{h_n})\beta + \Pi_{h_n}\beta - \beta.$$

   The tuning parameter $h_n$ is here given as $h_n = [n^{1/v_k}]$ where $v_k = 2a + 1 + \kappa_k$ for some small $\kappa_k > 0$ and $[a]$ denote the nearest integer of $a \in \mathbb{R}$. We also compute $t_{h_n}(X_0)$ and $\hat{t}_{h_n}(X_0)$. For further purposes, some quantities for prediction are also computed: $\hat{Y}_0 = \left\langle\hat{\beta}_{h_n}, X_0\right\rangle$ and $\hat{\varepsilon}_0 = Y_0 - \hat{Y}_0$.

3. Store the following quantities.

   - Variance term: $\Gamma^{\dagger}_{h_n}U_n$.
   - Random bias term 1: $(\hat{\Gamma}^{\dagger}_{h_n} - \Gamma^{\dagger}_{h_n})U_n$.
   - Random bias term 2: $(\hat{\Pi}_{h_n} - \Pi_{h_n})\beta$.
   - Non-random bias term: $\Pi_{h_n}\beta - \beta$.

- Truncated and non-truncated roots: $\hat{\beta}_{h_n} - \Pi_{h_n}\beta$ and $\hat{\beta}_{h_n} - \beta$.

- Scaling factors: $t_{h_n}(X_0)$, $\hat{t}_{h_n}(X_0)$ and $h_n$.

- $\hat{Y}_0$ and $\hat{\varepsilon}_0$.

Although we store all the quantities, for brevity, we only show the pictures for our main quantities $T_1 \equiv \sqrt{n/t_{h_n}(X_0)}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ and $T_2 \equiv \sqrt{n/h_n}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ in Figures 2.8-2.11.

### 2.10.3 Coverage rates when the new predictor is random

We will examine the empirical coverage rates and average widths for intervals when the new predictor $X_0$ is random, as these were not included in Section 2.5.2 of the main paper. At each Monte Carlo iteration, we simulate the new predictor $X_0$ (along with the corresponding error $\varepsilon_0$) as well as the data samples $\{(X_i, Y_i)\}_{i=1}^n$. Here, the Monte Carlo simulation size $M$ and the bootstrap sample size $Q$ are given as $M = 1000 = Q$. Figures 2.12-2.13 shows the empirical coverage rates and average widths for each interval when the new predictor $X_0$ is random.

The simulation are conducted as follows. For each $m = 1, \ldots, M$, perform the following.

1. (Simulation) Simulate $X_i$ with $X_i \overset{\mathrm{d}}{=} X$ and $\varepsilon_i \overset{\mathrm{d}}{=} \varepsilon$, where $\{(X_i, \varepsilon_i)\}_{i=1}^n$ are independent, and compute $Y_i = \langle\beta, X_i\rangle + \varepsilon_i$ for $i = 1, \ldots, n$. Also, for prediction purpose, simulate $X_0 \overset{\mathrm{d}}{=} X$ and $\varepsilon_0 \overset{\mathrm{d}}{=} \varepsilon$, where $(X_0, \varepsilon_0)$ is independent of $\{(X_i, \varepsilon_i)\}_{i=1}^n$, and compute $Y_0 = \langle\beta, X_0\rangle + \varepsilon_0$.

2. (Residuals) Compute the residuals $\hat{\varepsilon}_i = Y_i - \langle\hat{\beta}_{k_n}, X_i\rangle$ and the centered residuals $\tilde{\varepsilon}_i = \hat{\varepsilon}_i - \bar{\hat{\varepsilon}}$ for $i = 1, \ldots, n$.

3. (Residual bootstrap) To approximate the bootstrap distribution, do the following for each $q = 1, \ldots, Q$.

   (a) Draw independent bootstrap error $\{\varepsilon^*_{q,i}\}_{i=1}^n$ and $\varepsilon^*_{q,0}$ from the uniform distribution on the centered residuals $\{\tilde{\varepsilon}_i\}_{i=1}^n$.

   (b) Compute the bootstrap responses $Y^*_{q,i} = \langle\hat{\beta}_{g_n}, X_i\rangle + \varepsilon^*_{q,i}$, and construct the bootstrap estimate $\hat{\beta}^*_{q,h_n}$ based on the bootstrap samples $\{(X_i, Y^*_{q,i})\}_{i=1}^n$.

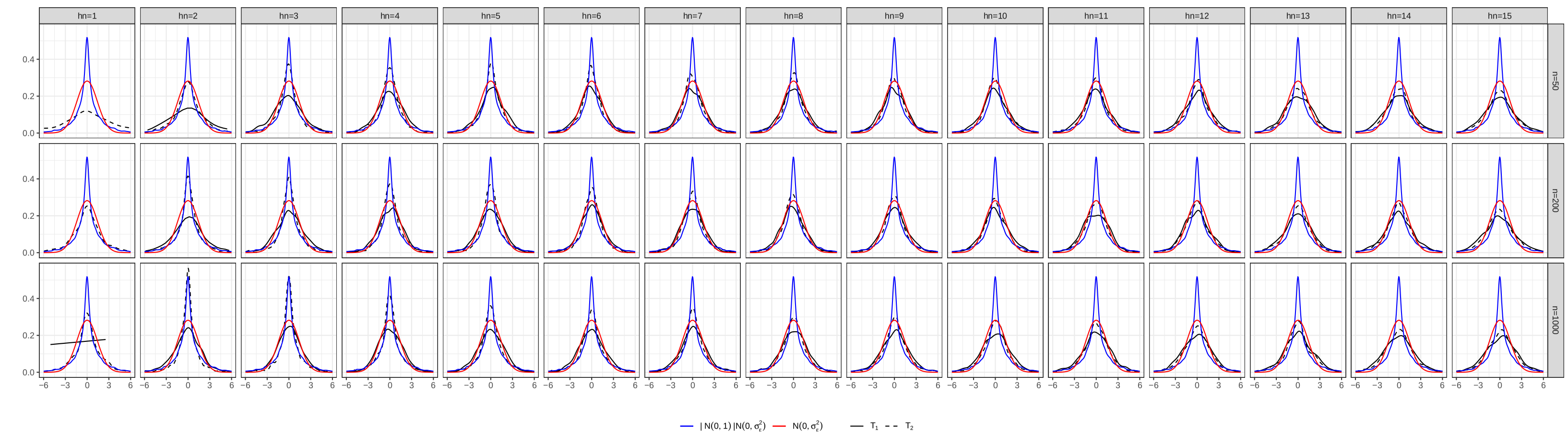


Figure 2.8: Kernel-estimated densities of $T_1 \equiv \{n/t_{h_n}(X_0)\}^{1/2}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ with scaling $t_{h_n}(X_0)$ (solid black line, according to our result) and $T_2 \equiv \{n/h_n\}^{1/2}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ with scaling $h_n$ (dashed black line, according to [CMS]) over different sample sizes $n$ and tuning parameters $h_n$ when $a = 2.5$ and $b = 2$. The theoretical limits $\mathsf{N}(0, \sigma_\varepsilon^2 = 2)$ of $T_1$ and $|\xi_0|Z_0$ of $T_2$ in Proposition 1 are given for reference (red and blue solid lines, respectively). The x-axis represents the values of $T_1$ and $T_2$ from -5 to 5 while the y-axis indicates the density values from 0 to 0.6. Each column and row respectively mean different tuning parameters $h_n \in \{1, \ldots, 15\}$ and sample sizes $n \in \{50, 200, 1000\}$.

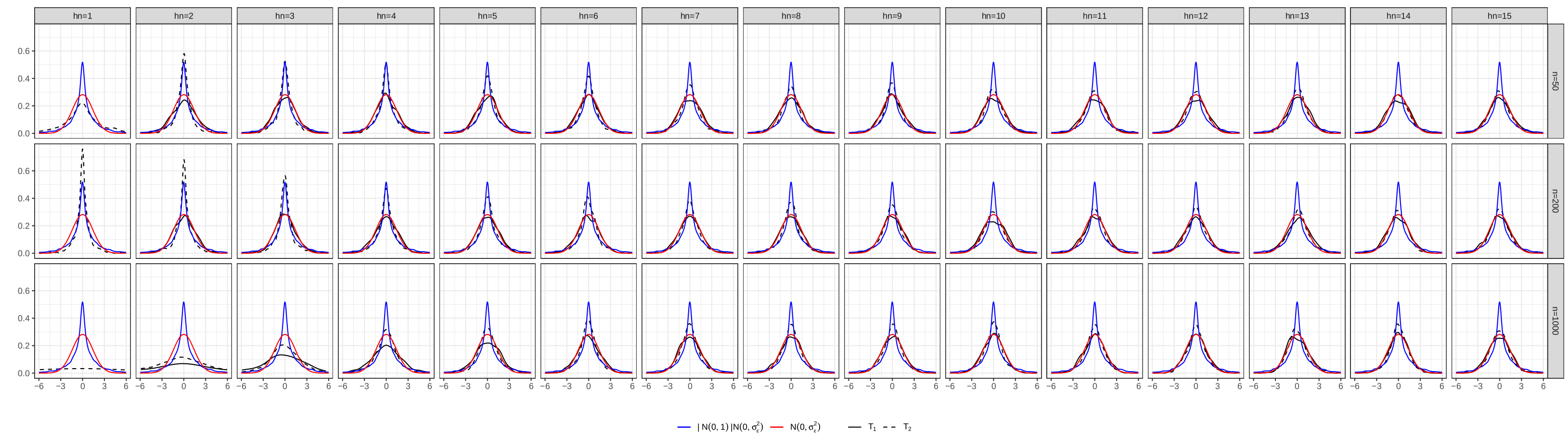


Figure 2.9: Kernel-estimated densities of $T_1 \equiv \{n/t_{h_n}(X_0)\}^{1/2}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ with scaling $t_{h_n}(X_0)$ (solid black line, according to our result) and $T_2 \equiv \{n/h_n\}^{1/2}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ with scaling $h_n$ (dashed black line, according to [CMS]) over different sample sizes $n$ and tuning parameters $h_n$ when $a = 2.5$ and $b = 5$. The theoretical limits $\mathsf{N}(0, \sigma_\varepsilon^2 = 2)$ of $T_1$ and $|\xi_0|Z_0$ of $T_2$ in Proposition 1 are given for reference (red and blue solid lines, respectively). The x-axis represents the values of $T_1$ and $T_2$ from -5 to 5 while the y-axis indicates the density values from 0 to 0.6. Each column and row respectively mean different tuning parameters $h_n \in \{1, \ldots, 15\}$ and sample sizes $n \in \{50, 200, 1000\}$.

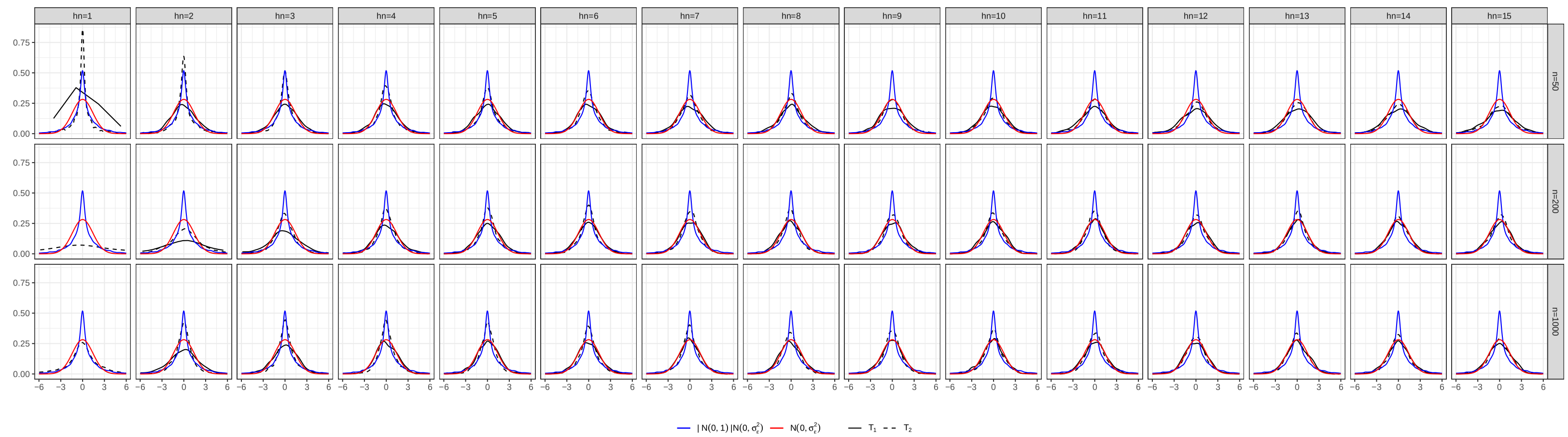


Figure 2.10: Kernel-estimated densities of $T_1 \equiv \{n/t_{h_n}(X_0)\}^{1/2}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ with scaling $t_{h_n}(X_0)$ (solid black line, according to our result) and $T_2 \equiv \{n/h_n\}^{1/2}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ with scaling $h_n$ (dashed black line, according to [CMS]) over different sample sizes $n$ and tuning parameters $h_n$ when $a = 5$ and $b = 2$. The theoretical limits $\mathsf{N}(0, \sigma_\varepsilon^2 = 2)$ of $T_1$ and $|\xi_0|Z_0$ of $T_2$ in Proposition 1 are given for reference (red and blue solid lines, respectively). The x-axis represents the values of $T_1$ and $T_2$ from -5 to 5 while the y-axis indicates the density values from 0 to 0.6. Each column and row respectively mean different tuning parameters $h_n \in \{1, \ldots, 15\}$ and sample sizes $n \in \{50, 200, 1000\}$.

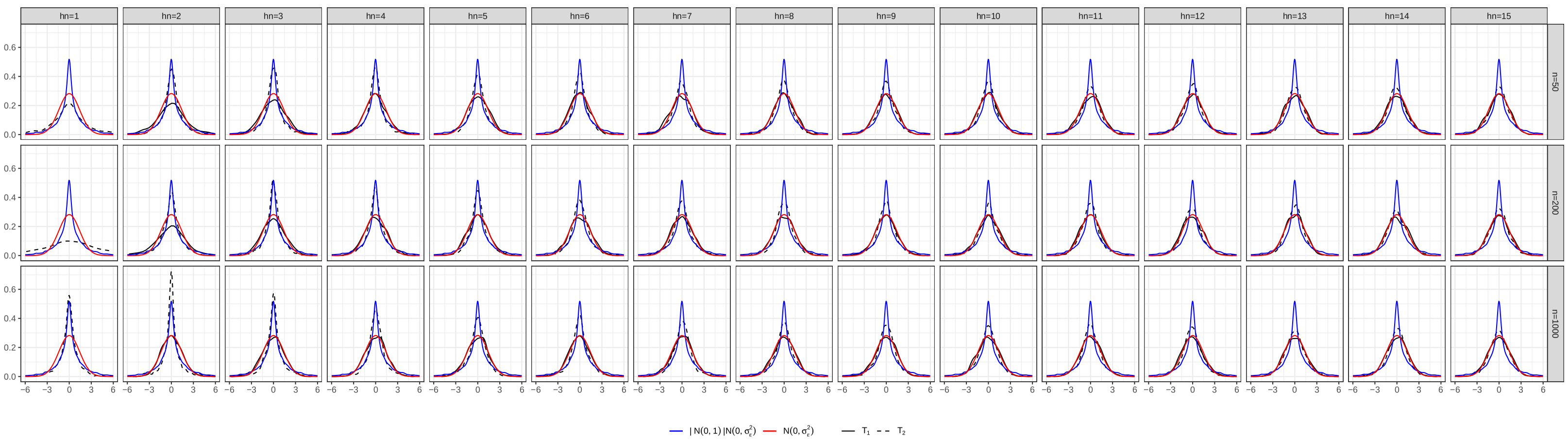


Figure 2.11: Kernel-estimated densities of $T_1 \equiv \{n/t_{h_n}(X_0)\}^{1/2}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ with scaling $t_{h_n}(X_0)$ (solid black line, according to our result) and $T_2 \equiv \{n/h_n\}^{1/2}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ with scaling $h_n$ (dashed black line, according to [CMS]) over different sample sizes $n$ and tuning parameters $h_n$ when $a = 5$ and $b = 5$. The theoretical limits $\mathsf{N}(0, \sigma_\varepsilon^2 = 2)$ of $T_1$ and $|\xi_0|Z_0$ of $T_2$ in Proposition 1 are given for reference (red and blue solid lines, respectively). The x-axis represents the values of $T_1$ and $T_2$ from -5 to 5 while the y-axis indicates the density values from 0 to 0.6. Each column and row respectively mean different tuning parameters $h_n \in \{1, \ldots, 15\}$ and sample sizes $n \in \{50, 200, 1000\}$.

(c) For prediction purporse, set $Y^*_{q,0} = \langle \hat{\beta}_{g_n}, X_0 \rangle + \varepsilon^*_{q,0}$ from the estimate $\hat{\beta}_{g_n}$, and compute the bootstrap prediction error $\hat{\varepsilon}^*_{q,0} = \hat{Y}^*_{q,0} - Y^*_{q,0}$ to approximate the prediction error $\hat{\varepsilon}_0 = \hat{Y}_0 - Y_0$ where $\hat{Y}^*_{q,0} = \langle \hat{\beta}^*_{q,h_n}, X_0 \rangle$ and $\hat{Y}_0 = \langle \hat{\beta}_{h_n}, X_0 \rangle$.

4. (Intervals based on the residual bootstrap)

For all cases, construct the following confidence intervals for $\langle \Pi_{h_n}\beta, X_0 \rangle$ and for $\langle \beta, X_0 \rangle$ and prediction intervals for $Y_0$ based on the residual bootstrap.

(a) Compute the $\alpha/2$ and $1-\alpha/2$ quantiles of $\{\langle \hat{\beta}^*_{q,h_n} - \hat{\Pi}_{h_n}\hat{\beta}_{g_n}, X_0 \rangle\}_{q=1}^{Q}$, say $l$ and $u$. Then, the confidence interval for $\langle \Pi_{h_n}\beta, X_0 \rangle$ is

$$CI^{trunc} = \left[\langle \hat{\beta}_{h_n}, X_0 \rangle - u, \langle \hat{\beta}_{h_n}, X_0 \rangle - l\right].$$

(b) Compute the $\alpha/2$ and $1-\alpha/2$ quantiles of $\{\langle \hat{\beta}^*_{q,h_n} - \hat{\beta}_{g_n}, X_{m,0} \rangle\}_{q=1}^{Q}$, say $l$ and $u$. Then, the confidence interval for $\langle \beta, X_0 \rangle$ is

$$CI = \left[\langle \hat{\beta}_{h_n}, X_0 \rangle - u, \langle \hat{\beta}_{h_n}, X_0 \rangle - l\right].$$

(c) Compute the $\alpha/2$ and $1-\alpha/2$ quantiles of $\{\hat{\varepsilon}^*_{q,0}\}_{q=1}^{Q}$, say $l$ and $u$. Then, the prediction interval for $Y_0$ is

$$PI = \left[\hat{Y}_0 - u, \hat{Y}_0 - l\right].$$

5. (Intervals based on the central limit theorem) For all cases, construct the following confidence intervals for $\langle \Pi_{h_n}\beta, X_0 \rangle$ and for $\langle \beta, X_0 \rangle$ and prediction intervals for $Y_0$ based on central limit theorem.

(a) The confidence interval for $\langle \Pi_{h_n}\beta, X_0 \rangle$ is

$$CI^{trunc} = \left[\langle \hat{\beta}_{h_n}, X_0 \rangle - \hat{\sigma}_\varepsilon z_{1-\alpha/2}\sqrt{\frac{\hat{t}_{h_n}(X_0)}{n}}, \langle \hat{\beta}_{h_n}, X_0 \rangle + \hat{\sigma}_\varepsilon z_{1-\alpha/2}\sqrt{\frac{\hat{t}_{h_n}(X_0)}{n}}\right].$$

(b) The confidence interval for $\langle \beta, X_{m,0} \rangle$ is

$$CI = \left[\langle \hat{\beta}_{h_n}, X_0 \rangle - \hat{\sigma}_\varepsilon z_{1-\alpha/2}\sqrt{\frac{\hat{t}_{h_n}(X_0)}{n}}, \langle \hat{\beta}_{h_n}, X_0 \rangle + \hat{\sigma}_\varepsilon z_{1-\alpha/2}\sqrt{\frac{\hat{t}_{h_n}(X_0)}{n}}\right].$$

(c) The prediction interval for $Y_0$ is

$$PI = \left[\hat{Y}_0 - \hat{\sigma}_\varepsilon z_{1-\alpha/2}\sqrt{\frac{\hat{t}_{h_n}(X_0)}{n} + 1}, \hat{Y}_0 + \hat{\sigma}_\varepsilon z_{1-\alpha/2}\sqrt{\frac{\hat{t}_{h_n}(X_0)}{n} + 1}\right].$$

6. Let $T_0$ denote the target quantity, either $\langle \Pi_{h_n}\beta, X_0\rangle$, $\langle \beta, X_0\rangle$, or $Y_0$, and $I$ denote the corresponding interval constructed above, i.e., either $CI^{trunc}$, $CI$, or $PI$, for each method (one of the residual bootstrap and the central limit theorem). Compute $I_m = \mathbb{I}(T_0 \in I)$.

The coverage probability $1 - \alpha$ is then approximated by $M^{-1}\sum_{m=1}^{M} I_m$.

Figures 2.12-2.13 respectively provide an illustration of empirical coverage rates and average widths for each interval with different tuning parameters $h_n$ and $g_n$ under the scenarios considered.

### 2.10.4 Coverage rates when the new predictor is fixed

In addition to the results in Section 2.5.3 of the main paper, we provide further coverage probabilities and average widths for the following intervals when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed.

- Individual confidence intervals (ICIs) for $\langle \Pi_{h_n}\beta, X_0\rangle$ and $\langle \beta, X_0\rangle$ based on residual bootstrap and central limit theorem. These are referred to as ICI_trunc and ICI, respectively.

- Individual prediction intervals (IPIs) for $Y_0 = \langle \beta, X_0\rangle + \varepsilon_0$ based on residual bootstrap and central limit theorem. This is referred to as IPI.

- Simultaneous confidence intervals (SCIs) for $\langle \Pi_{h_n}\beta, X_0\rangle$ and $\langle \beta, X_0\rangle$ based on residual bootstrap either with or without studentization. These are referred to as SCI_trunc and SCI without studentization and SCI_trunc_std and SCI_std with studentization, respectively.

- Simultaneous prediction intervals (SPIs) for $Y_0 = \langle \beta, X_0\rangle + \varepsilon_0$ based on residual bootstrap either with or without studentization. These are SPI_trunc_std and SPI_std with studentization, respectively.

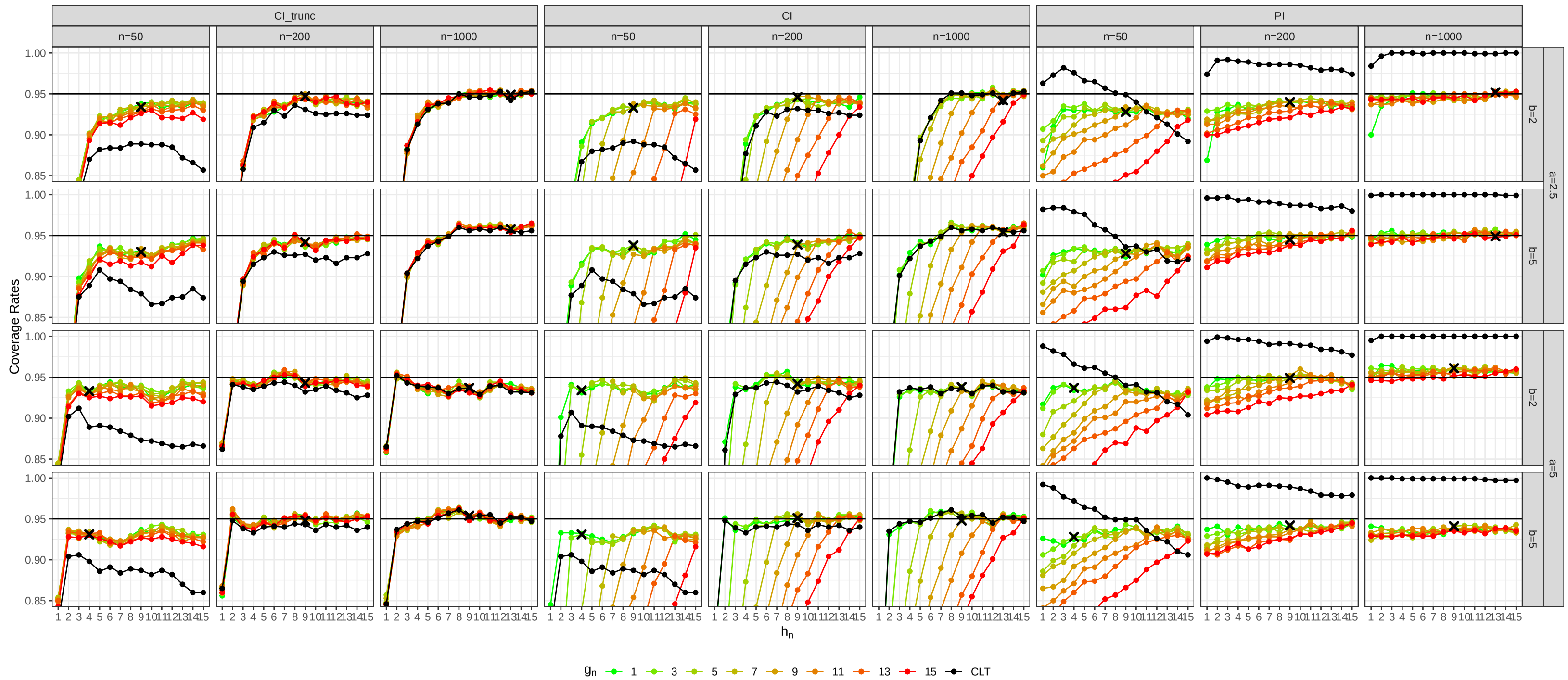


Figure 2.12: Empirical coverages of intervals from bootstrap and normal approximation over different truncations when the new predictor $X_0$ is random. The three columns display the CI for $\langle \Pi_{h_n}\beta, X_0\rangle$, the CI for $\langle \beta, X_0\rangle$, and the PI for $Y_0$, respectively. Crosses $\times$ indicate bootstrap coverages with $h_n, g_n$ selected by a proposed rule. The sub-columns in each column represents different sample sizes $n \in \{50, 200, 1000\}$. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

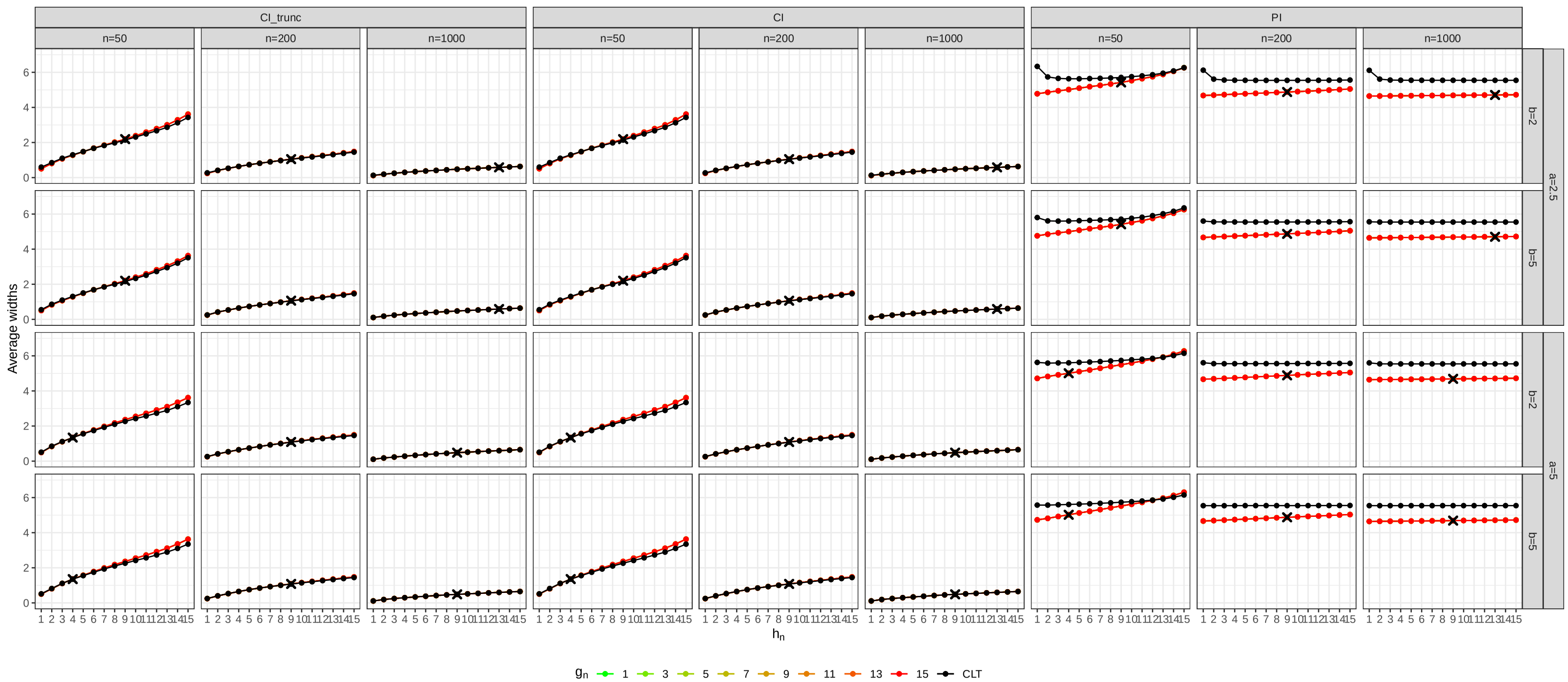


Figure 2.13: Average widths of intervals from bootstrap and normal approximation over different truncations when the new predictor $X_0$ is random. Crosses $\times$ indicate bootstrap coverages with $h_n, g_n$ selected by the proposed rule. The sub-columns in each column represents different sample sizes $n \in \{50, 200, 1000\}$. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

As seen in the below construction, the SCI_std and SPI are constructed based on Corollaries 3-4 in the main paper. However, in practice, it turns out from our simulation that the studentization does not have a substantial effect in terms of empirical coverage rates and average widths.

In contrast to the case when the new predictor is random, here we simulate the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and fix it before the Monte Carlo iteration. We now describe the detailed procedure of simulation. The Monte Carlo simulation size $M$ and the bootstrap sample size $Q$ are again provided as $M = 1000 = Q$. Given the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$, for each $m = 1, \ldots, M$, perform the following.

1. (Simulation) Simulate $X_i$ with $X_i \overset{\mathrm{d}}{=} X$ and $\varepsilon_i \overset{\mathrm{d}}{=} \varepsilon$, where $\{(X_i, \varepsilon_i)\}_{i=1}^{n}$ are independent, and compute $Y_i = \langle \beta, X_i \rangle + \varepsilon_i$ for $i = 1, \ldots, n$. Also, for prediction purpose, simulate independent $\varepsilon_{0,i_0} \overset{\mathrm{d}}{=} \varepsilon$ for $i_0 = 1, \ldots, n_0$, where $\{\varepsilon_{0,i_0}\}_{i_0=1}^{n_0}$ is independent of $\{(X_i, \varepsilon_i)\}_{i=1}^{n}$, and compute $Y_{0,i_0} = \langle \beta, X_{0,i_0} \rangle + \varepsilon_{0,i_0}$ for $i_0 = 1, \ldots, n_0$.

2. (Residuals) Compute the residuals $\hat{\varepsilon}_i = Y_i - \langle \hat{\beta}_{k_n}, X_i \rangle$ and the centered residuals $\tilde{\varepsilon}_i = \hat{\varepsilon}_i - \bar{\hat{\varepsilon}}$ for $i = 1, \ldots, n$.

3. (Residual bootstrap) For $q = 1, \ldots, Q$, do the following.

   (a) Draw independent bootstrap error $\varepsilon^*_{q,i}$ for $i = 1, \ldots, n$ and $\varepsilon^*_{q,0,i_0}$ from the uniform distribution on the centered residuals $\{\tilde{\varepsilon}_i\}_{i=1}^{n}$

   (b) Compute the bootstrap responses $Y^*_{q,i} = \langle \hat{\beta}_{g_n}, X_i \rangle + \varepsilon^*_{q,i}$ from the estimate $\hat{\beta}_{g_n}$, and construct the bootstrap estimate $\hat{\beta}^*_{q,h_n}$ based on the bootstrap samples $\{(X_i, Y^*_{q,i})\}_{i=1}^{n}$.

   (c) For prediction purpose, for each $i_0 = 1, \ldots, n_0$, set $Y^*_{q,0,i_0} = \langle \hat{\beta}_{g_n}, X_{0,i_0} \rangle + \varepsilon^*_{q,0,i_0}$, and compute the bootstrap prediction error $\hat{\varepsilon}^*_{q,0,i_0} = \hat{Y}^*_{q,0,i_0} - Y^*_{q,0,i_0}$ to approximate the prediction error $\hat{\varepsilon}_{0,i_0} = \hat{Y}_{0,i_0} - Y_{0,i_0}$ where $\hat{Y}^*_{q,0,i_0} = \langle \hat{\beta}^*_{q,h_n}, X_{0,i_0} \rangle$ and $\hat{Y}_{0,i_0} = \langle \hat{\beta}_{h_n}, X_{0,i_0} \rangle$.

4. (Intervals based on the residual bootstrap) Construct the following intervals based on the residual bootstrap.

   (a) ICIs for $\{\langle \Pi_{h_n} \beta, X_{0,i_0} \rangle\}_{i_0=1}^{n_0}$.

Compute the $100(1-\alpha/2)\%$ and $100(\alpha/2)\%$ quantiles of $\{\langle\hat{\beta}^*_{q,h_n} - \hat{\Pi}_{q,h_n}\hat{\beta}_{g_n}, X_{0,i_0}\rangle\}_{q=1}^{Q}$, say $u_{i_0}$ and $l_{i_0}$. Then, the ICIs for $\{\langle\Pi_{h_n}\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ are

$$ICI_{i_0}^{trunc} = \left[\langle\hat{\beta}_{h_n}, X_{0,i_0}\rangle - u_{i_0}, \langle\hat{\beta}_{h_n}, X_{0,i_0}\rangle - l_{i_0}\right].$$

(b) ICIs for $\{\langle\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$.

Compute the $100(1-\alpha/2)\%$ and $100(\alpha/2)\%$ quantiles of $\{\langle\hat{\beta}^*_{q,h_n} - \hat{\beta}_{g_n}, X_{0,i_0}\rangle\}_{q=1}^{Q}$, say $u_{i_0}$ and $l_{i_0}$. Then, the ICIs for $\{\langle\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ are

$$ICI_{i_0} = \left[\langle\hat{\beta}_{h_n}, X_{0,i_0}\rangle - u_{i_0}, \langle\hat{\beta}_{h_n}, X_{0,i_0}\rangle - l_{i_0}\right].$$

(c) SCIs for $\{\langle\Pi_{h_n}\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$.

Compute the $100(1-\alpha/2)\%$ quantile of $\{\max_{1\le i_0\le n_0}|\langle\hat{\beta}^*_{q,h_n} - \hat{\Pi}_{h_n}\hat{\beta}_{g_n}, X_{0,i_0}\rangle|\}_{q=1}^{Q}$, say $u$. Then, the SCIs for $\{\langle\Pi_{h_n}\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ are

$$SCI_{i_0}^{trunc} = \left[\langle\hat{\beta}_{h_n}, X_0\rangle - u, \langle\hat{\beta}_{h_n}, X_0\rangle + u\right].$$

(d) SCIs for $\{\langle\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$.

Compute the $100(1-\alpha/2)\%$ quantile of $\{\max_{1\le i_0\le n_0}|\langle\hat{\beta}^*_{q,h_n} - \hat{\beta}_{g_n}, X_{0,i_0}\rangle|\}_{q=1}^{Q}$, say $u$. Then, the SCIs for $\{\langle\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ are

$$SCI_{i_0} = \left[\langle\hat{\beta}_{h_n}, X_0\rangle - u, \langle\hat{\beta}_{h_n}, X_0\rangle + u\right].$$

(e) Studentized SCIs for $\{\langle\Pi_{h_n}\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$.

Compute the $100(1-\alpha/2)\%$ quantile of

$$\left\{\max_{1\le i_0\le n_0}\frac{|\langle\hat{\beta}^*_{q,h_n} - \hat{\Pi}_{h_n}\hat{\beta}_{g_n}, X_{0,i_0}\rangle|}{\sqrt{\hat{t}_{h_n}(X_{0,i_0})}}\right\}_{q=1}^{Q},$$

say $u$. Then, the studentized SCIs for $\{\langle\Pi_{h_n}\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ are

$$SCI_{i_0}^{trunc,std} = \left[\langle\hat{\beta}_{h_n}, X_0\rangle - \sqrt{\hat{t}_{h_n}(X_{0,i_0})}u, \langle\hat{\beta}_{h_n}, X_0\rangle + \sqrt{\hat{t}_{h_n}(X_{0,i_0})}u\right].$$

(f) Studentized SCIs for $\{\langle\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$.

Compute the $100(1-\alpha/2)\%$ quantile of

$$\left\{ \max_{1\le i_0 \le n_0} \frac{\left|\langle \hat{\beta}^*_{q,h_n} - \hat{\beta}_{g_n}, X_{0,i_0}\rangle\right|}{\sqrt{\hat{t}_{h_n}(X_{0,i_0})}} \right\}_{q=1}^{Q},$$

say $u$. Then, the studentized SCIs for $\{\langle \beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ are

$$SCI^{std}_{i_0} = \left[ \langle \hat{\beta}_{h_n}, X_0\rangle - \sqrt{\hat{t}_{h_n}(X_{0,i_0})}u, \langle \hat{\beta}_{h_n}, X_0\rangle + \sqrt{\hat{t}_{h_n}(X_{0,i_0})}u \right].$$

(g) IPIs for $\{Y_{0,i_0}\}_{i_0=1}^{n_0}$.

Compute the $100(1-\alpha/2)\%$ and $100(\alpha/2)\%$ quantiles of $\{\hat{\varepsilon}^*_{q,0,i_0}\}_{q=1}^{Q}$, say $u_{i_0}$ and $l_{i_0}$. Then, the IPIs for $\{Y_{0,i_0}\}_{i_0=1}^{n_0}$ are

$$IPI_{i_0} = [\hat{Y}_{0,i_0} - u_{i_0}, \hat{Y}_{0,i_0} - l_{i_0}].$$

(h) SPIs for $\{Y_{0,i_0}\}_{i_0=1}^{n_0}$

Compute the $100(1-\alpha/2)\%$ quantile of $\left\{ \max_{1\le i_0\le n_0} \left| Y^*_{0,i_0,q} - \hat{Y}^*_{0,i_0,q} \right| \right\}_{q=1}^{Q}$, say $u$. Then, the SPIs for $\{Y_{0,i_0}\}_{i_0=1}^{n_0}$ are

$$SPI_{i_0} = [\hat{Y}_{0,i_0} - u, \hat{Y}_{0,i_0} + u].$$

(i) Studentized SPIs for $\{Y_{0,i_0}\}_{i_0=1}^{n_0}$

Compute the $100(1-\alpha/2)\%$ quantile of

$$\left\{ \max_{1\le i_0 \le n_0} \frac{\left| Y^*_{0,i_0,q} - \hat{Y}^*_{0,i_0,q} \right|}{\sqrt{\hat{t}_{h_n}(X_{0,i_0})/n + 1}} \right\}_{q=1}^{Q},$$

say $u$. Then, the studentized SPIs for $\{Y_{0,i_0}\}_{i_0=1}^{n_0}$ are

$$SPI^{std}_{m} = \left[ \hat{Y}_{0,i_0} - \sqrt{\hat{t}_{h_n}(X_{0,i_0})/n + 1}u, \hat{Y}_{0,i_0} + \sqrt{\hat{t}_{h_n}(X_{0,i_0})/n + 1}u \right].$$

5. (Intervals based on central limit theorem) For each case, construct the following intervals based on central limit theorem.

(a) ICIs for $\{\langle \Pi_{h_n}\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$:

$$ICI_{i_0}^{trunc} = \left[\langle \hat{\beta}_{h_n}, X_{0,i_0}\rangle - z_{1-\alpha/2}\sqrt{\frac{\hat{t}_{h_n}(X_{0,i_0})}{n}}, \langle \hat{\beta}_{h_n}, X_{0,i_0}\rangle + z_{1-\alpha/2}\sqrt{\frac{\hat{t}_{h_n}(X_{0,i_0})}{n}}\right].$$

(b) ICIs for $\{\langle \beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$:

$$ICI_{i_0} = \left[\langle \hat{\beta}_{h_n}, X_{0,i_0}\rangle - z_{1-\alpha/2}\sqrt{\frac{\hat{t}_{h_n}(X_{0,i_0})}{n}}, \langle \hat{\beta}_{h_n}, X_{0,i_0}\rangle + z_{1-\alpha/2}\sqrt{\frac{\hat{t}_{h_n}(X_{0,i_0})}{n}}\right].$$

(c) IPIs for $\{Y_{0,i_0}\}_{i_0=1}^{n_0}$:

$$IPI_{i_0} = \left[\hat{Y}_{0,i_0} - z_{1-\alpha/2}\sqrt{\frac{\hat{t}_{h_n}(X_{0,i_0})}{n} + 1}, \hat{Y}_{0,i_0} + z_{1-\alpha/2}\sqrt{\frac{\hat{t}_{h_n}(X_{0,i_0})}{n} + 1}\right].$$

6. Let $\{T_{0,i_0}\}_{i_0=1}^{n_0}$ denote the target quantities, either $\{\langle \Pi_{h_n}\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$, $\{\langle \beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$, or $\{Y_{0,i_0}\}_{i_0=1}^{n_0}$. Also, $\{II_{i_0}\}_{i_0=1}^{n_0}$ denote the corresponding individual intervals constructed above, i.e., either $\{ICI_{i_0}^{trunc}\}_{i_0=1}^{n_0}$, $\{ICI_{i_0}\}_{i_0=1}^{n_0}$, or $\{IPI_{i_0}\}_{i_0=1}^{n_0}$, for each method (one of the residual bootstrap and the central limit theorem). Finally, $\{SI_{i_0}\}_{i_0=1}^{n_0}$ denote the corresponding simultaneous intervals constructed above, that are either studentized or not, i.e., either $\{SCI_{i_0}^{trunc}\}_{i_0=1}^{n_0}$, $\{SCI_{i_0}\}_{i_0=1}^{n_0}$, $\{SCI_{i_0}^{trunc,std}\}_{i_0=1}^{n_0}$, or $\{SCI_{i_0}^{std}\}_{i_0=1}^{n_0}$.

The coverage probabilities for individual intervals are approximated by

$$1 - \hat{\alpha}_{i_0} = M^{-1}\sum_{m=1}^{M} \mathbb{I}(T_{0,i_0} \in II_{i_0})$$

for each $i_0 = 1, \ldots, n_0$, and those for simultaneous intervals are approximately computed as

$$1 - \hat{\alpha} = M^{-1}\sum_{m=1}^{M} \mathbb{I}(T_{0,i_0} \in SI_{i_0}, \forall i_0 = 1, \ldots, n_0).$$

We have the results for two sets of new predictors. One consists of the first five eigenfunctions $\mathbb{V}_1 = \{e_1, \ldots, e_5\}$. Figures 2.14-2.16 show the empirical coverage rates of ICIs for $\{\langle \Pi_{h_n}\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$, ICIs for $\{\langle \beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$, and IPIs for $\{Y_{0,i_0}\}_{i_0=1}^{n_0}$, respectively, while the corresponding average widths are displayed in Figures 2.17-2.18. Note that the widths of ICIs for the truncated and original projections are equal due to their construction based on residual bootstrap. Empirical coverage rates of SCIs for $\{\langle \Pi_{h_n}\beta, X_{0,i_0}\}_{i_0=1}^{n_0}$, SCIs for $\{\langle \beta, X_{0,i_0}\}_{i_0=1}^{n_0}$, and SPIs for $\{Y_{0,i_0}\}_{i_0=1}^{n_0}$ are provided in Figure 2.19. In Figure 2.20, we illustrate the average widths of non-studentized SCIs and SPIs. As these are not based on studentization, the widths do not depend on each of the new predictors. Meanwhile, the average wdiths of studentized SCIs and SPIs are given in Figures 2.21-2.22 since studentization causes the widths to depend on the new predictors. As above, the widths of studentized SCIs for the truncated and original projections are equal. In the other set $\mathbb{V}_2$, the new predictors $X_{0,i_0}$ are independently drawn from the model $X \overset{\mathrm{d}}{=} \sum_{j=1}^{J} \sqrt{\lambda_j}\xi_j e_j$ introduced in Section 2.5.1 of the main paper. Figures 2.23-2.31 shows the results for $\mathbb{V}_2$ where the panels are arranged in the same order as the figures for $\mathbb{V}_1$.

## 2.11 Additional details regarding real data analysis

In this section, we provide further background for real data analysis on wheat moisture data from Section 2.6 of the main paper.

### 2.11.1 Selection of the tuning parameter $k_n$ by using a cross-validation method based on prediction error

When we conduct the residual bootstrap procedure in practice, one can choose the tuning parameter $k_n$ for determining residuals by a cross-validation method based on a certain measure such as the prediction error. We refer to [1] for a general overview of cross-validation methods. In our real data analysis, the procedure of selecting $k_n$ is as follows. For a given $k_n$ and each $m = 1, \ldots, M$, perform the following.

1. Divide the samples into training and testing samples with sizes $n_{tr}$ and $n_{test}$, respectively, with $n_{tr} + n_{test} = n$. Write $\mathcal{I}_{tr}$ and $\mathcal{I}_{test}$ for the corresponding index sets.

2. Compute the estimator $\hat{\beta}_{k_n,tr}$ with the tuning parameter $k_n$ based on the training sample $\{(X_i, Y_i) : i \in \mathcal{I}_{tr}\}$.

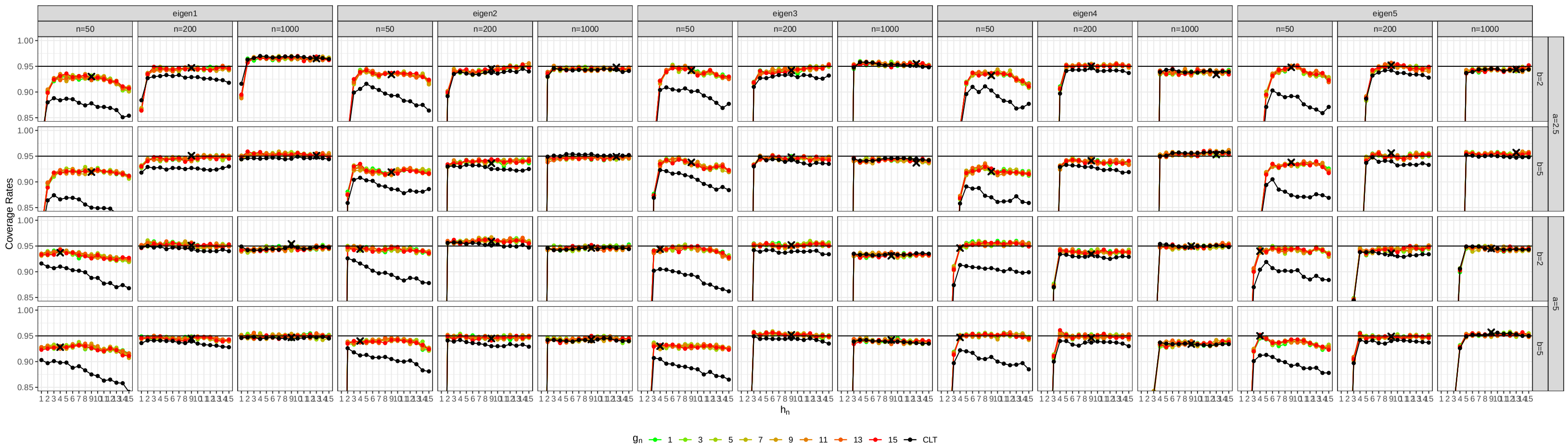


Figure 2.14: Empirical coverages of ICIs for $\{\langle \Pi_{h_n}\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ from bootstrap and normal approximation over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0} = e_{i_0}$. Crosses $\times$ indicate bootstrap coverages with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

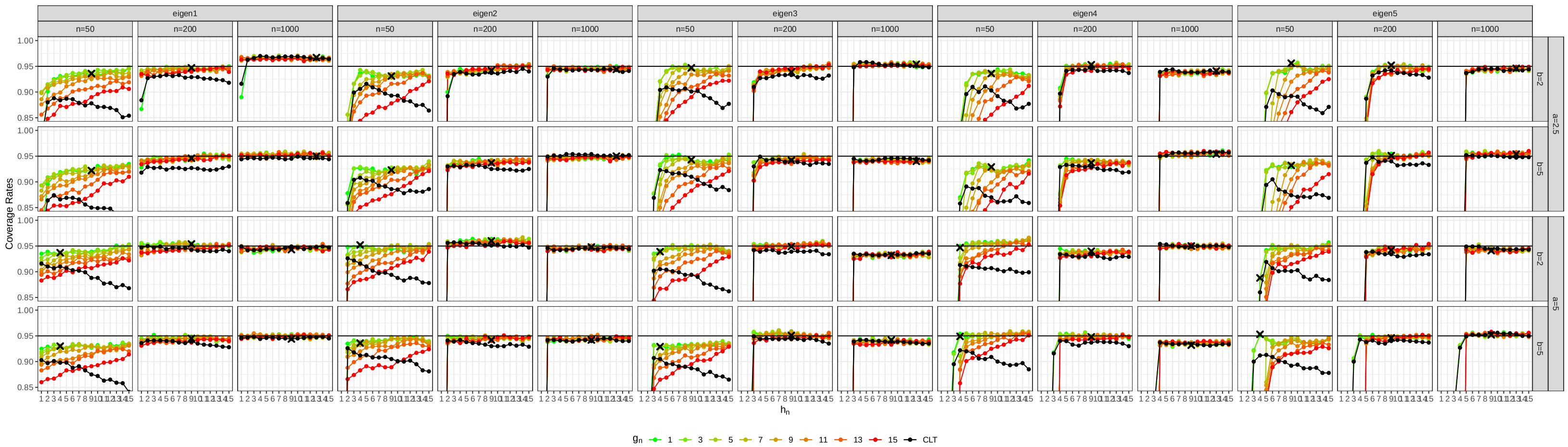


Figure 2.15: Empirical coverages of ICIs for $\{\langle\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ from bootstrap and normal approximation over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0} = e_{i_0}$. Crosses $\times$ indicate bootstrap coverages with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

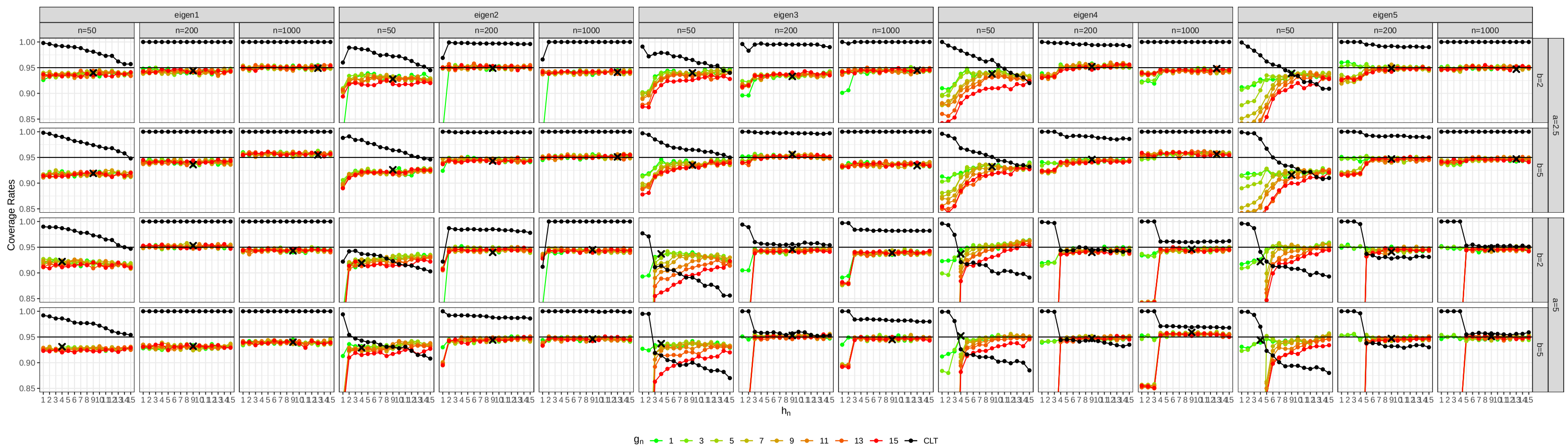


Figure 2.16: Empirical coverages of IPIs for $\{Y_{0,i_0}\}_{i_0=1}^{n_0}$ from bootstrap and normal approximation over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0} = e_{i_0}$. Crosses $\times$ indicate bootstrap coverages with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

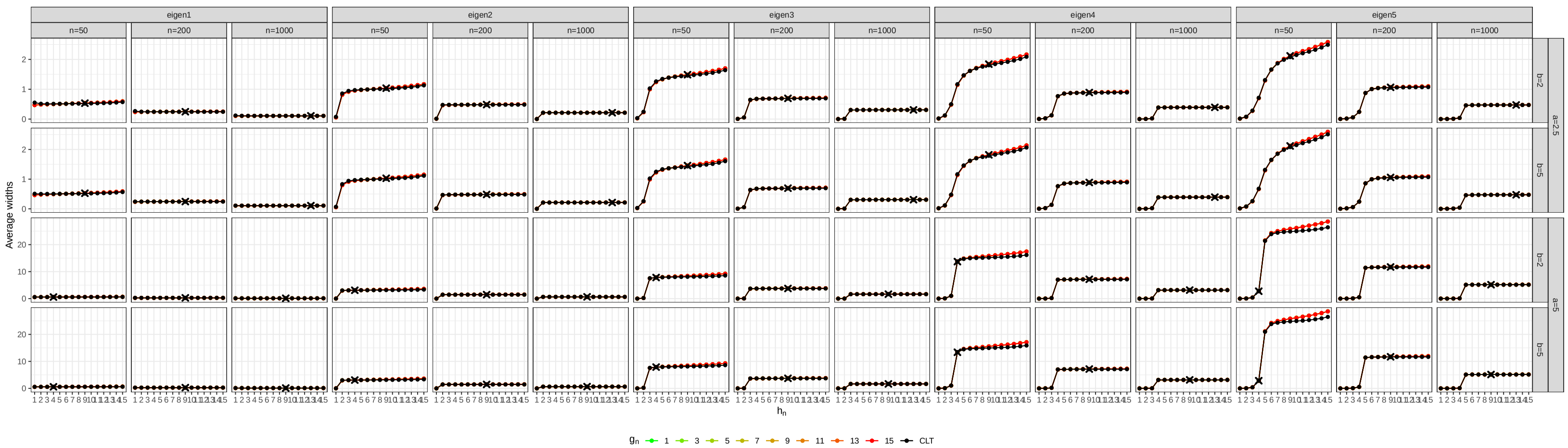


Figure 2.17: Average widths of ICIs for $\{\langle \Pi_{h_n}\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ and $\{\langle \beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ from bootstrap and normal approximation over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0} = e_{i_0}$. Crosses $\times$ indicate average widths with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

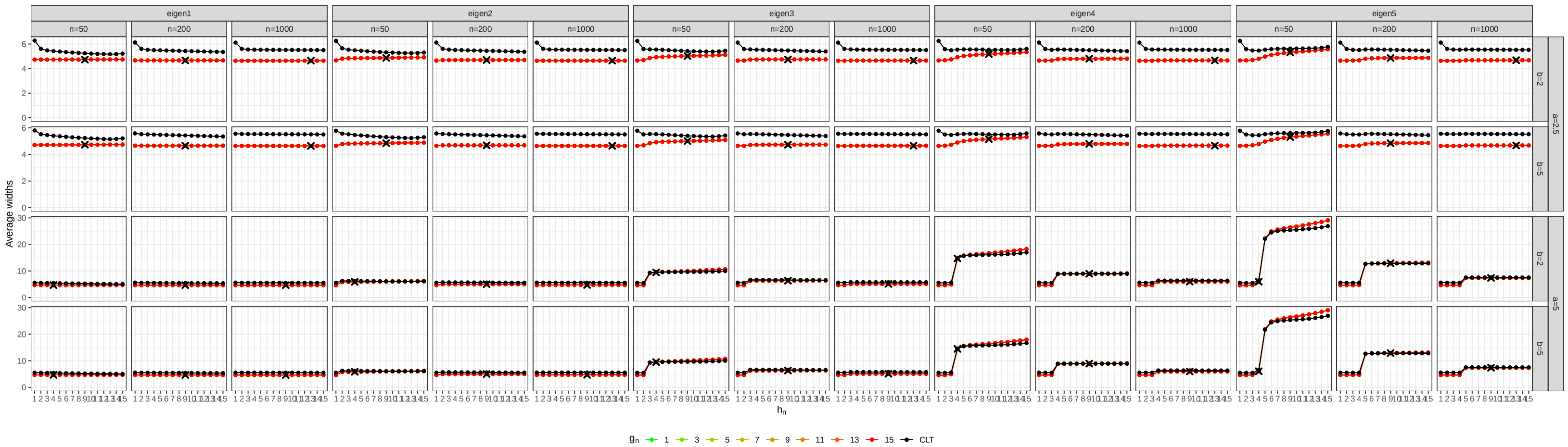


Figure 2.18: Average widths of IPIs for $\{Y_{0,i_0}\}_{i_0=1}^{n_0}$ from bootstrap and normal approximation over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0} = e_{i_0}$. Crosses $\times$ indicate average widths with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

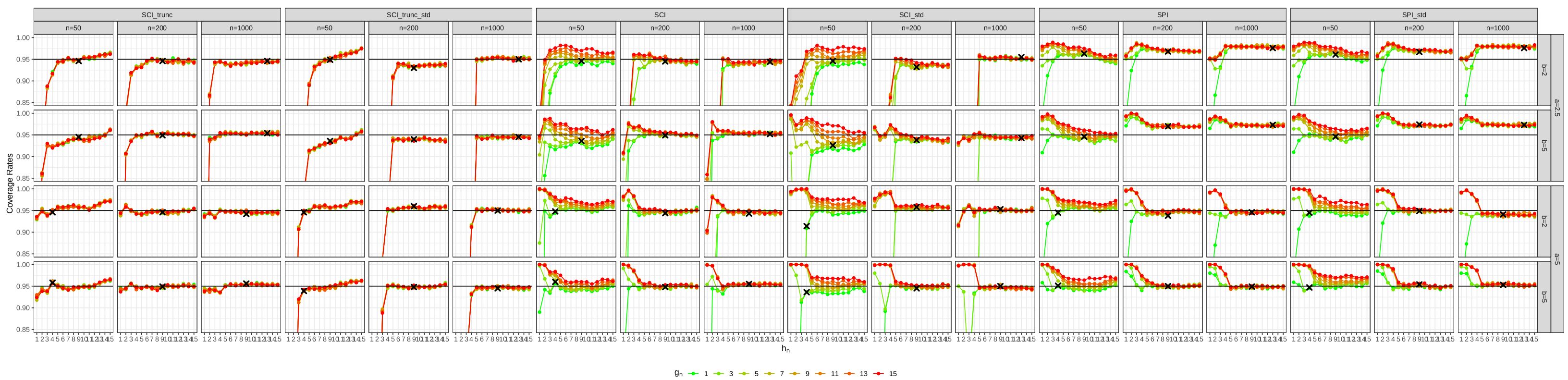


Figure 2.19: Empirical coverages of SCIs and SPIs from bootstrap over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0} = e_{i_0}$. Crosses $\times$ indicate bootstrap coverages with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

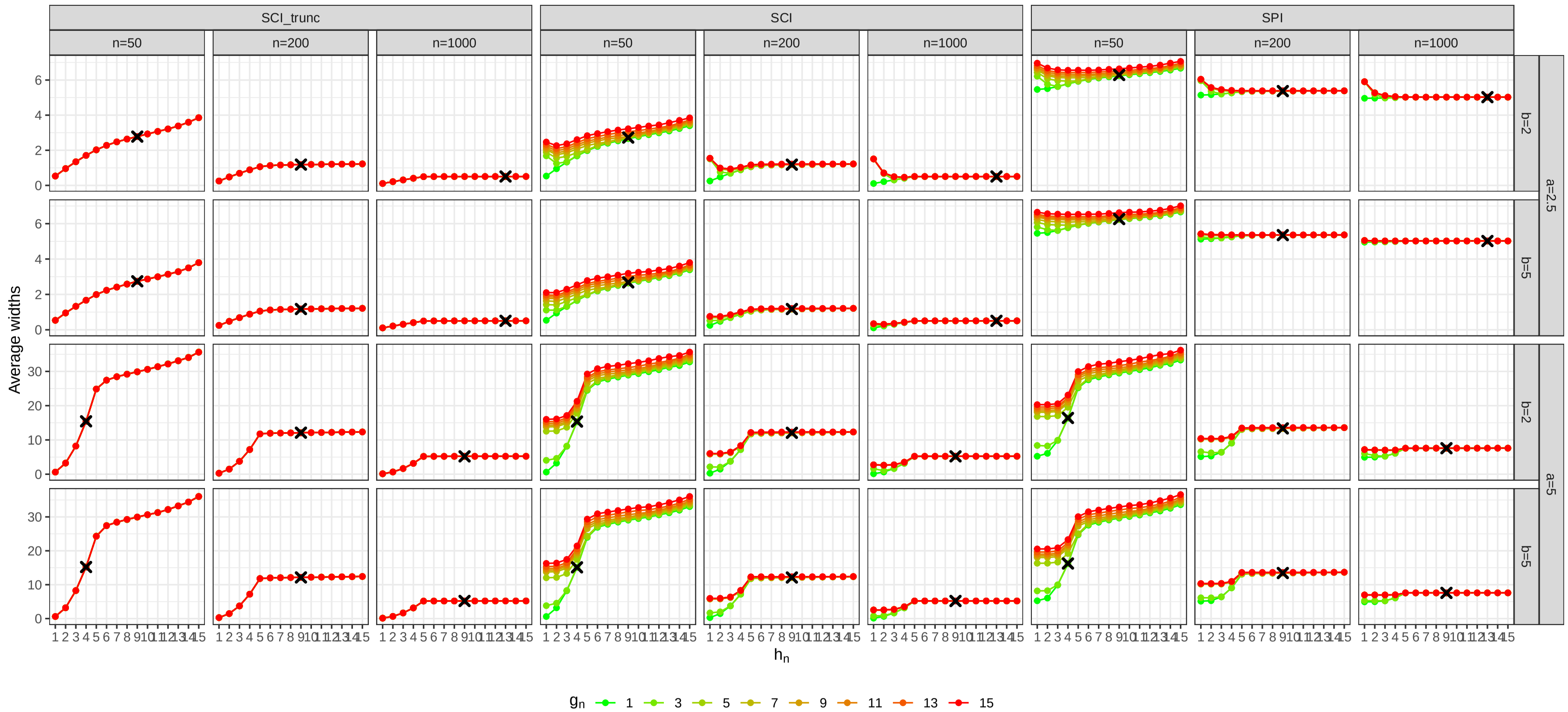


Figure 2.20: Average widths of non-studentized SCIs and SPIs from bootstrap over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0} = e_{i_0}$. Crosses $\times$ indicate average widths with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

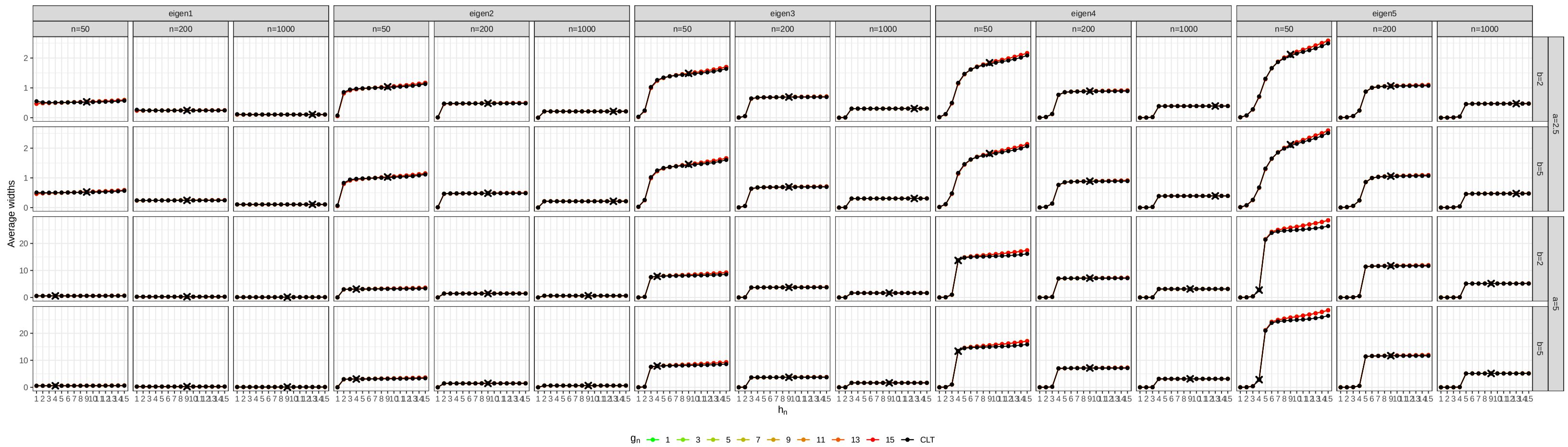


Figure 2.21: Average widths of studentized SCIs for $\{\langle \Pi_{h_n}\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ and $\{\langle \beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ from bootstrap over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0} = e_{i_0}$. Crosses $\times$ indicate average widths with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

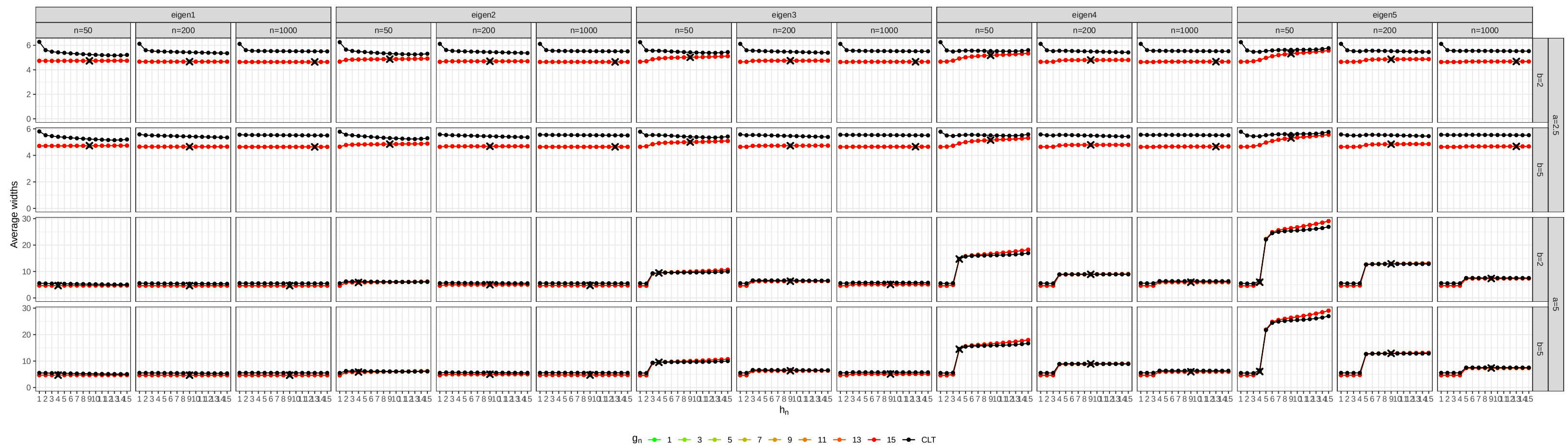


Figure 2.22: Average widths of studentized SPIs for $\{Y_{0,i_0}\}_{i_0=1}^{n_0}$ from bootstrap over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0} = e_{i_0}$. Crosses $\times$ indicate average widths with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

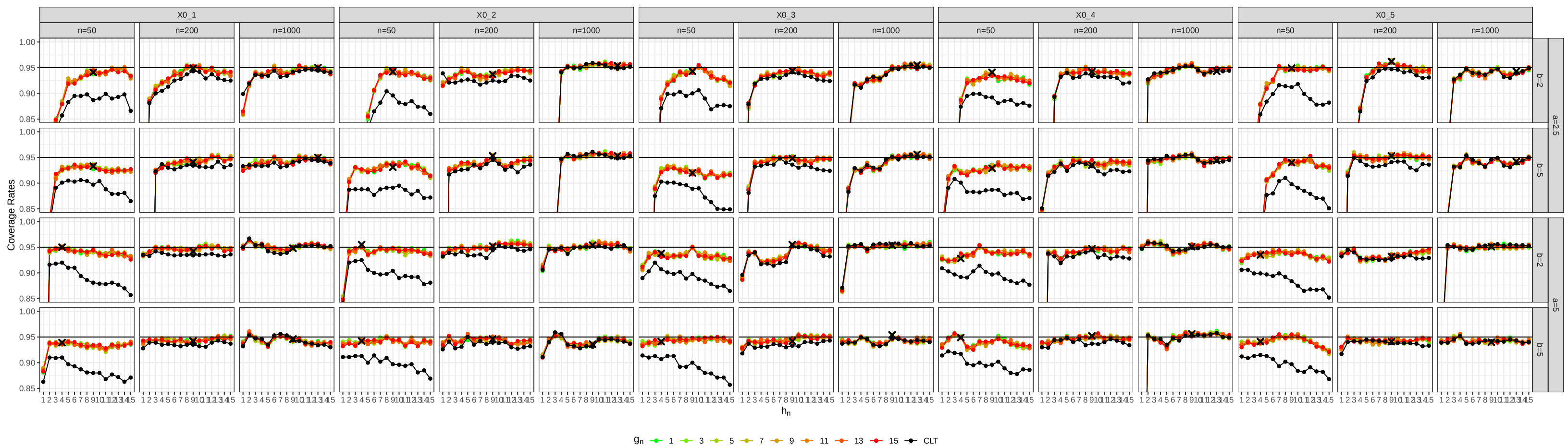


Figure 2.23: Empirical coverages of ICIs for $\{\langle \Pi_{h_n}\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ from bootstrap and normal approximation over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0}$ being drawn from the distribution of $X$. Crosses $\times$ indicate bootstrap coverages with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

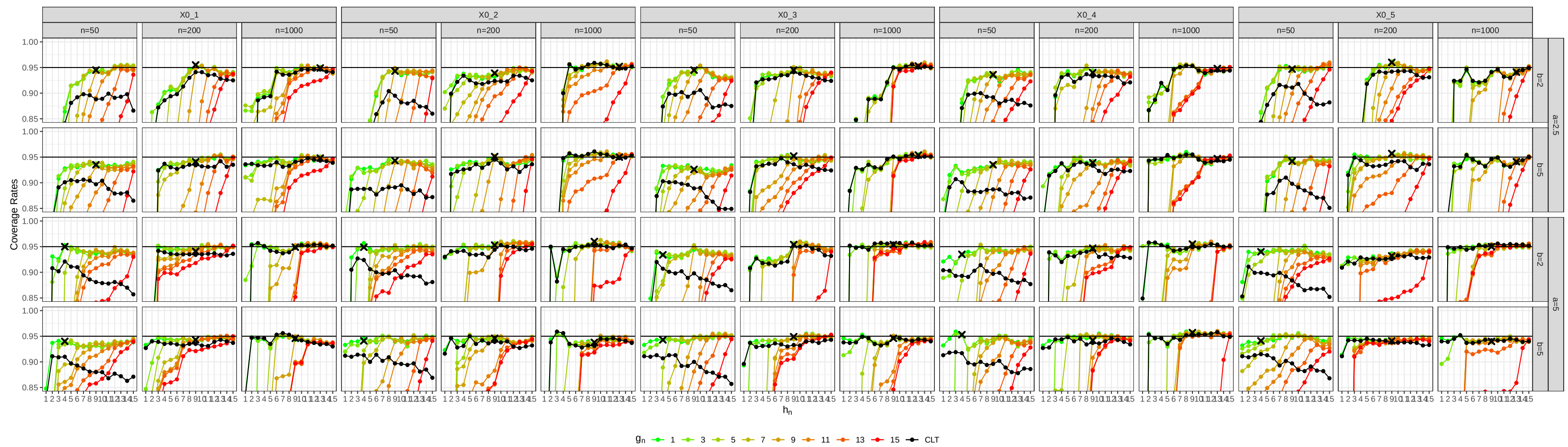


Figure 2.24: Empirical coverages of ICIs for $\{\langle \beta, X_{0,i_0} \rangle\}_{i_0=1}^{n_0}$ from bootstrap and normal approximation over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0}$ being drawn from the distribution of $X$. Crosses $\times$ indicate bootstrap coverages with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

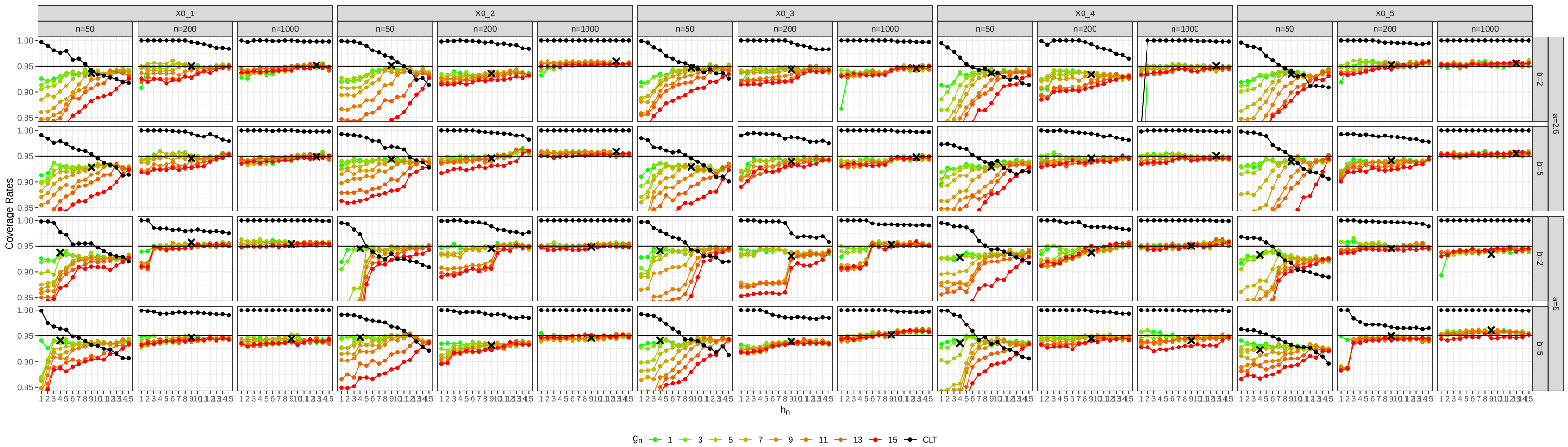


Figure 2.25: Empirical coverages of IPIs for $\{Y_{0,i_0}\}_{i_0=1}^{n_0}$ from bootstrap and normal approximation over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0}$ being drawn from the distribution of $X$. Crosses $\times$ indicate bootstrap coverages with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

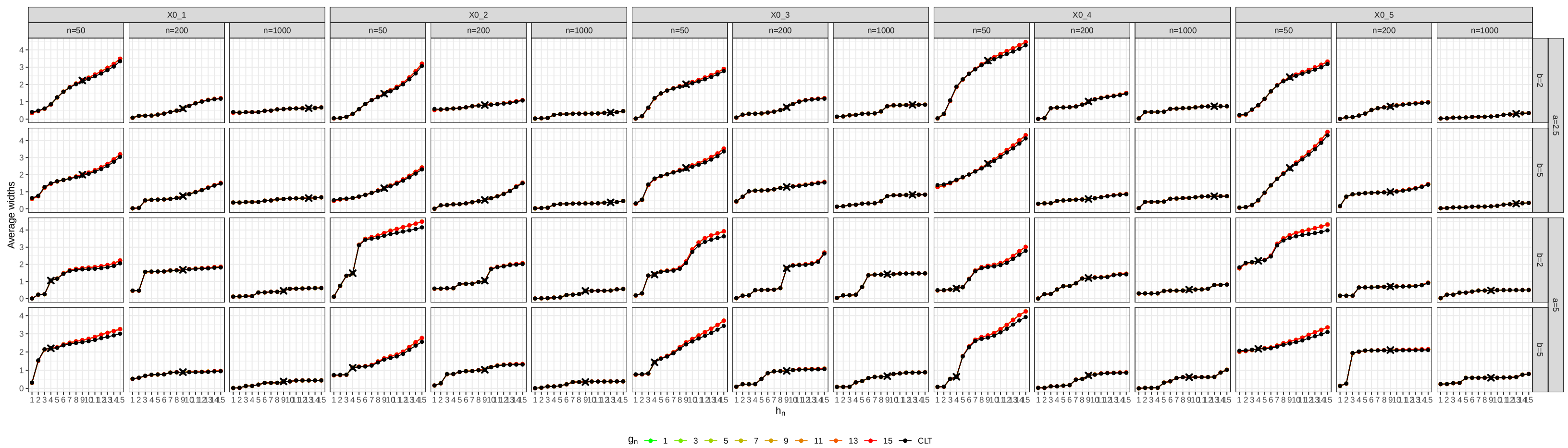


Figure 2.26: Average widths of ICIs for $\{\langle \Pi_{h_n}\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ and $\{\langle \beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ from bootstrap and normal approximation over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0}$ being drawn from the distribution of $X$. Crosses $\times$ indicate average widths with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \dots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \dots, 13, 15\}$ while the black line stands for the normal approximation.

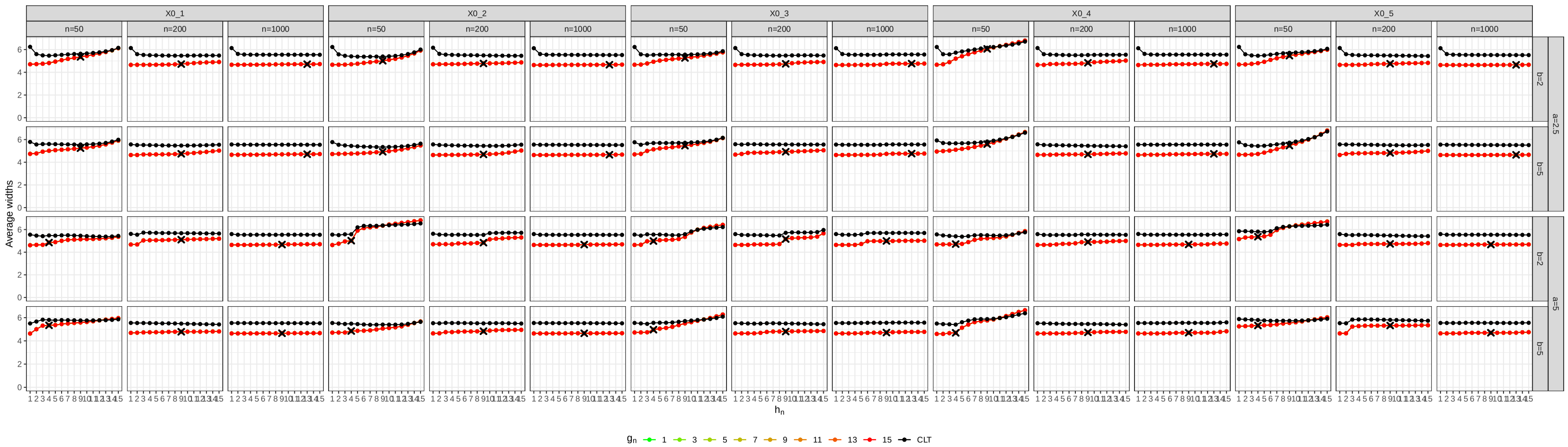


Figure 2.27: Average widths of IPIs for $\{Y_{0,i_0}\}_{i_0=1}^{n_0}$ from bootstrap and normal approximation over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0}$ being drawn from the distribution of $X$. Crosses $\times$ indicate average widths with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

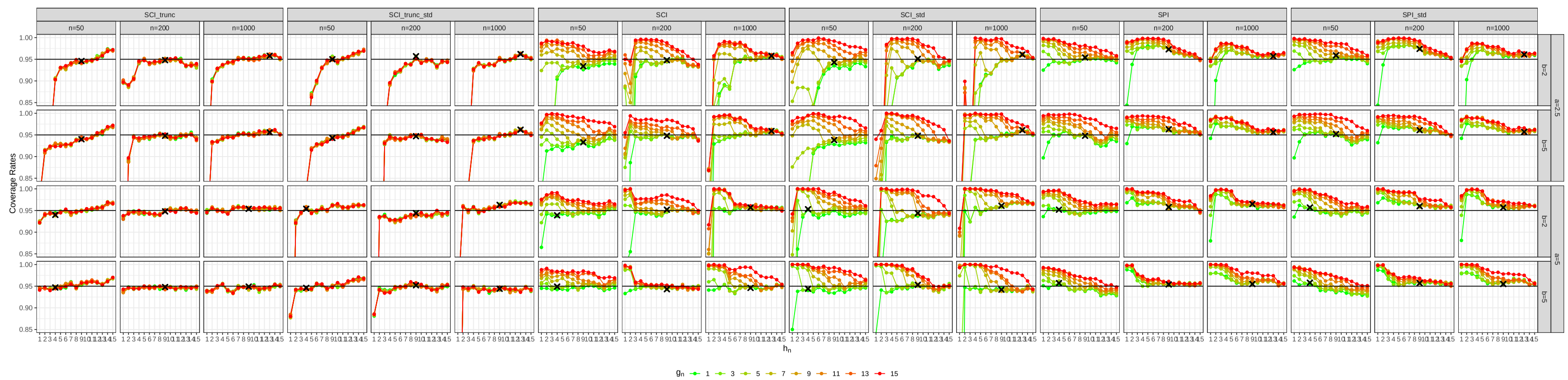


Figure 2.28: Empirical coverages of SCIs and SPIs from bootstrap over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0}$ being drawn from the distribution of $X$. Crosses $\times$ indicate bootstrap coverages with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

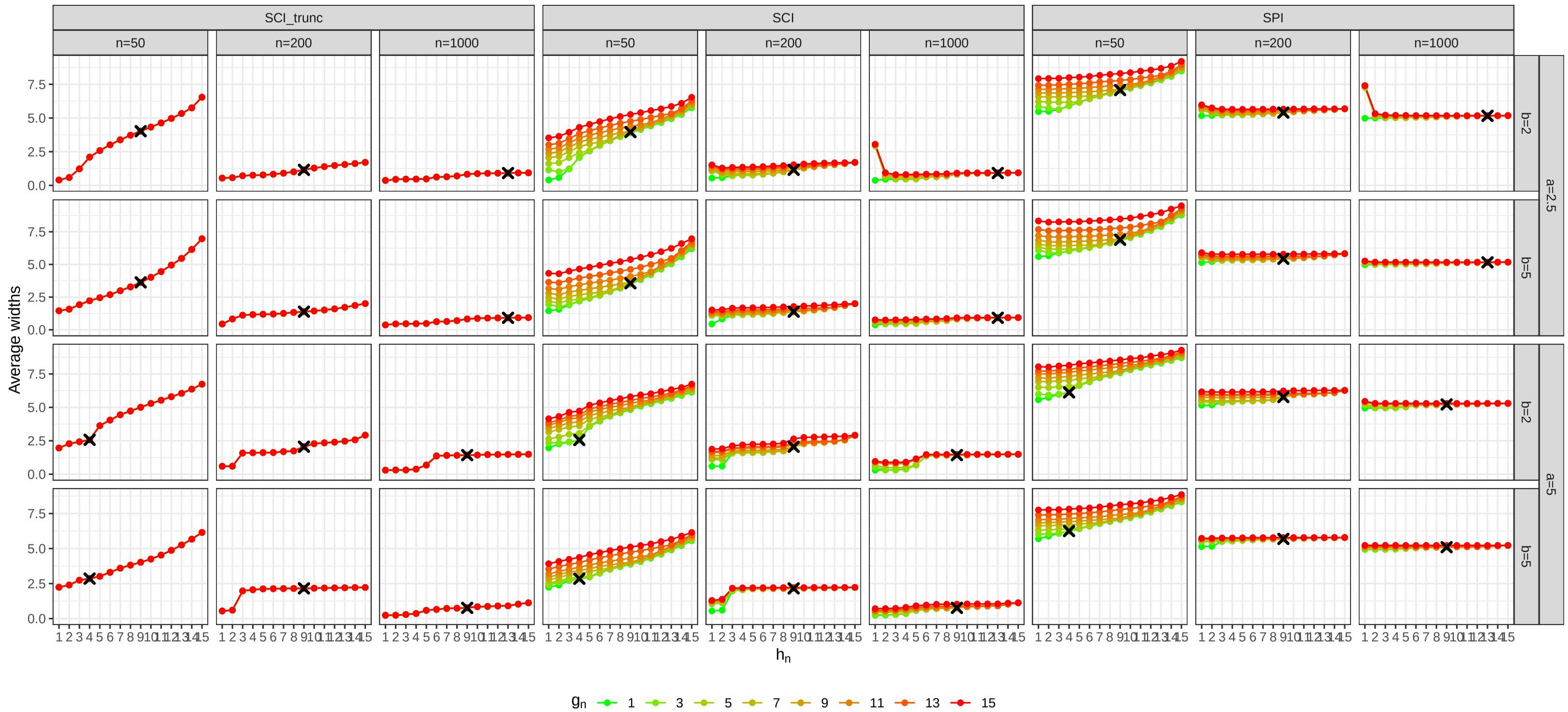


Figure 2.29: Average widths of non-studentized SCIs and SPIs from bootstrap over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0}$ being drawn from the distribution of $X$. Crosses $\times$ indicate average widths with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

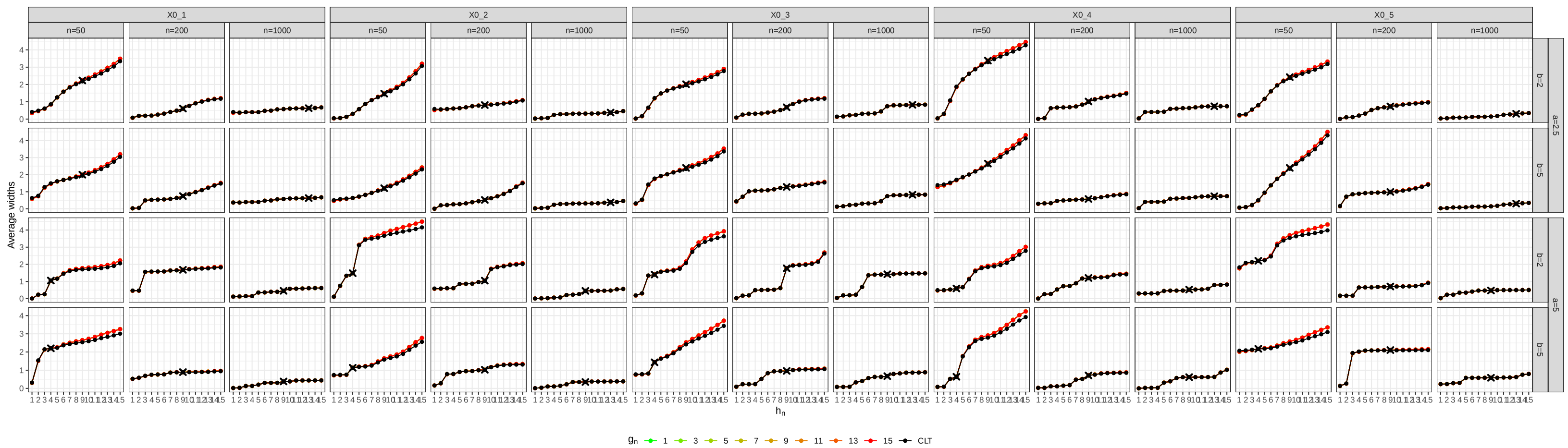


Figure 2.30: Average widths of studentized SCIs for $\{\langle \Pi_{h_n}\beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ and $\{\langle \beta, X_{0,i_0}\rangle\}_{i_0=1}^{n_0}$ from bootstrap over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0}$ being drawn from the distribution of $X$. Crosses $\times$ indicate average widths with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

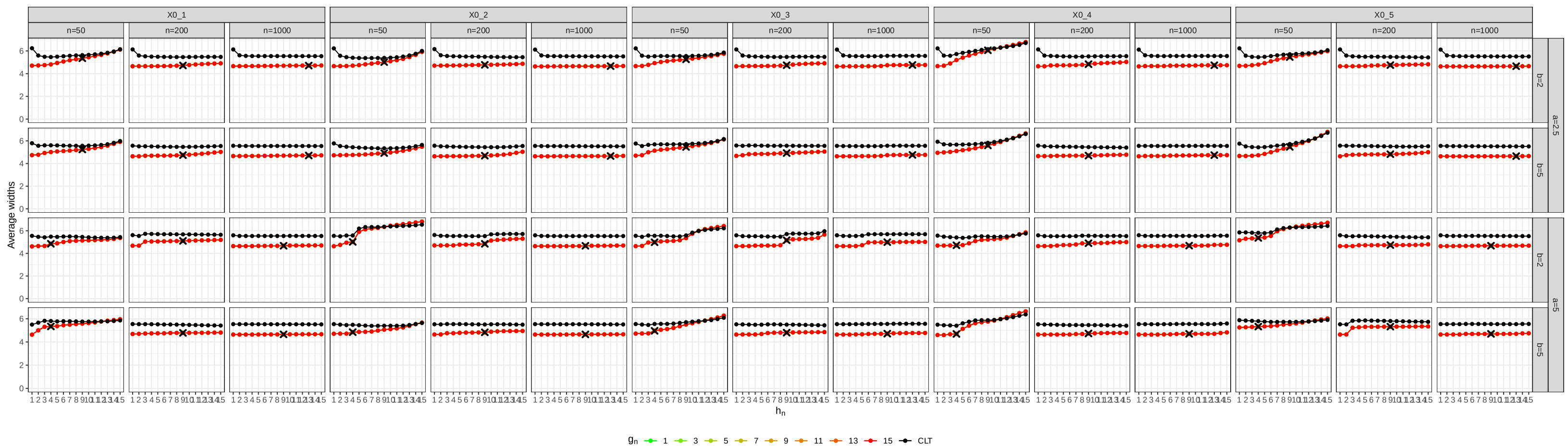


Figure 2.31: Average widths of studentized SPIs for $\{Y_{0,i_0}\}_{i_0=1}^{n_0}$ from bootstrap over different truncations when the new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ are fixed with $X_{0,i_0}$ being drawn from the distribution of $X$. Crosses $\times$ indicate average widths with $h_n, g_n$ selected by a proposed rule. The main columns and sub-columns represent different new predictors $\{X_{0,i_0}\}_{i_0=1}^{n_0}$ and different sample sizes $n \in \{50, 200, 1000\}$, respectively. The y-axis is the coverage rates for each interval and the x-axis denotes the tuning parameters $h_n \in \{1, \ldots, 15\}$. Different colors from green to red mean the other tuning parameters $g_n \in \{1, 3, \ldots, 13, 15\}$ while the black line stands for the normal approximation.

3. Compute the prediction error $PE_m(k_n) = n_{test}^{-1} \sum_{i \in \mathcal{I}_{test}} (Y_i - \langle \hat{\beta}_{k_n,tr}, X_i \rangle)^2$ of $\hat{\beta}_{k_n,tr}$ based on the testing sample $\{(X_i, Y_i) : i \in \mathcal{I}_{test}\}$.

We now compute the estimate $\widehat{PE}(k_n)$ of the true prediction error as

$$\widehat{PE}(k_n) \equiv M^{-1} \sum_{m=1}^{M} PE_m(k_n).$$

Among some pilot tuning parameters (e.g., $\{1, \ldots, 20\}$) for $k_n$, one can choose one that minimizes the estimated prediction error $\widehat{PE}(k_n)$.

### 2.11.2 Symmetrized intervals based on the residual bootstrap

One may construct a symmetrized version of individual confidence or prediction intervals based on the residual bootstrap. For simplicity, we describe this only with ICI for untrucated projection $\langle \beta, X_0 \rangle$ and IPI for the new response $Y_0$. We first supppose that we have the bootstrap samples, which provides the bootstrap estimators $\{\hat{\beta}^*_{h_n,q}\}_{q=1}^{Q}$ and the bootstrap new response $\{Y^*_{q,0}\}_{q=1}^{Q}$, where $Q$ denotes the Monte Carlo size to approximate the bootstrap distribution. Then, the symmetrized ICI/IPI are obtained as follows.

1. (ICI) Compute the $100(1-\alpha/2)\%$ quantiles of $\{|\langle \hat{\beta}^*_{h_n,q} - \hat{\beta}_{g_n}, X_0 \rangle|\}_{q=1}^{Q}$, say $u$. Then, the symmetrized confidence interval for $\langle \beta, X_0 \rangle$ is

$$ICI = \left[ \langle \hat{\beta}_{h_n}, X_0 \rangle - u, \langle \hat{\beta}_{h_n}, X_0 \rangle + u \right].$$

2. (IPI) Compute the $100(1-\alpha/2)\%$ quantiles of $\{|\hat{Y}^*_{q,0} - Y^*_{q,0}\}_{q=1}^{Q}$, say $u$, where $\hat{Y}^*_{q,0} \equiv \langle \hat{\beta}^*_{h_n,q}, X_0 \rangle$. Then, the symmetrized confidence interval for $Y_0$ is

$$IPI = \left[ \langle \hat{\beta}_{h_n}, X_0 \rangle - u, \langle \hat{\beta}_{h_n}, X_0 \rangle + u \right].$$

### 2.11.3 New predictor functions under consideration

The following functions are the new regressor functions that are considered in the real data analysis as described in Figure 2.7 of the main paper.

1. Overall shift: $X_{OS,1}(t) = 0.25$, $X_{OS,2}(t) = 0.15$, $X_{OS,3}(t) = 0.05$, $X_{OS,4}(t) = -0.05$, $X_{OS,5}(t) = -0.15$, $X_{OS,6}(t) = -0.25$

2. Simiple functions: write $\mathbb{I}_1(t) = \mathbb{I}(t \leq 1400)$, $\mathbb{I}_2(t) = \mathbb{I}(1400 < t \leq 1900)$, $\mathbb{I}_3(t) = \mathbb{I}(t > 1900)$, and define $X_{sim,1} = 0.1 \cdot \mathbb{I}_1$, $X_{sim,2} = 0.1 \cdot \mathbb{I}_2$, $X_{sim,3} = 0.1 \cdot \mathbb{I}_3$, $X_{sim,4} = -0.1 \cdot \mathbb{I}_1$, $X_{sim,5} = -0.1 \cdot \mathbb{I}_2$, $X_{sim,6} = -0.1 \cdot \mathbb{I}_3$.

3. Sums of two simple functions:

   - $X_{SS,1} = 0.1 \cdot \mathbb{I}_1 + 0.05 \cdot \mathbb{I}_2$, $X_{SS,2} = 0.1 \cdot \mathbb{I}_2 + 0.05 \cdot \mathbb{I}_3$, $X_{SS,3} = 0.1 \cdot \mathbb{I}_3 + 0.05 \cdot \mathbb{I}_1$,
   - $X_{SS,4} = -X_{SS,1}$, $X_{SS,5} = -X_{SS,2}$, $X_{SS,6} = -X_{SS,3}$.

### 2.11.4 Estimated slop function for the wheat data

In Figure 2.32, we provide an illustration of the estimated slope functions used in bootstrap inference with different tuning parameters $k_n = 4$, $h_n = 9$, and $g_n = 5$.

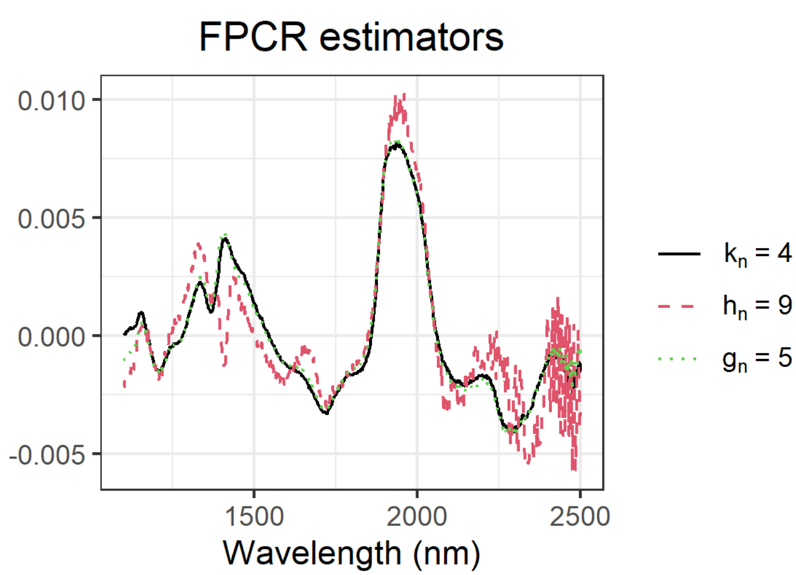


Figure 2.32: Estimated slope functions with the selected tuning parameters $k_n = 4$, $h_n = 9$, and $g_n = 5$

## 2.12 References

[1] Arlot, S., and Celisse, A. A survey of cross-validation procedures for model selection. *Statistics surveys 4* (2010), 40–79.

[2] Athreya, K. B., and Lahiri, S. N. *Measure Theory and Probability Theory.* Springer Science & Business Media, 2006.

[3] Bickel, P. J., and Freedman, D. A. Some asymptotic theory for the bootstrap. *The Annals of Statistics* (1981), 1196–1217.

[4] Billingsley, P. *Probability and Measure.* John Wiley & Sons, 1999.

[5] Cai, T. T., and Hall, P. Prediction in functional linear regression. *The Annals of Statistics 34*, 5 (2006), 2159–2179.

[6] Cardot, H., Ferraty, F., and Sarda, P. Functional linear model. *Statistics & Probability Letters 45*, 1 (1999), 11–22.

[7] Cardot, H., Mas, A., and Sarda, P. Clt in functional linear regression models. *Probability Theory and Related Fields 138*, 3-4 (2007), 325–361.

[8] Crambes, C., and Mas, A. Asymptotics of prediction in functional linear regression with functional outputs. *Bernoulli 19*, 5B (2013), 2627–2651.

[9] Crambes, C., and Mas, A. Asymptotics of prediction in functional linear regression with functional outputs. *Bernoulli 19*, 5B (2013), 2627–2651.

[10] Dezeure, R., Bühlmann, P., and Zhang, C.-H. High-dimensional simultaneous inference with the bootstrap. *Test 26*, 4 (2017), 685–719.

[11] Dunford, N., and Schwartz, J. T. *Linear Operators, Part 1: General Theory*, vol. 10. John Wiley & Sons, 1988.

[12] Ferraty, F., and Vieu, P. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Science & Business Media, 2006.

[13] Franke, J., and Hardle, W. On bootstrapping kernel spectral estimates. *The Annals of Statistics* (1992), 121–145.

[14] Freedman, D. A. Bootstrapping regression models. *The Annals of Statistics 9*, 6 (1981), 1218–1228.

[15] Gervini, D. Dynamic retrospective regression for functional data. *Technometrics 57*, 1 (Jan. 2015), 26–34. Publisher: Taylor & Francis.

[16] González-Manteiga, W., and Martínez-Calvo, A. Bootstrap in functional linear regression. *Journal of Statistical Planning and Inference 141*, 1 (2011), 453–461.

[17] Hall, P. Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *The Annals of Statistics* (1992a), 675–694.

[18] Hall, P. On bootstrap confidence intervals in nonparametric regression. *The Annals of Statistics* (1992b), 695–711.

[19] Hall, P., and Horowitz, J. L. Methodology and convergence rates for functional linear regression. *The Annals of Statistics 35*, 1 (2007), 70–91.

[20] Horváth, L., and Kokoszka, P. *Inference for Functional Data with Applications*, vol. 200. Springer Science & Business Media, 2012.

[21] Hsing, T., and Eubank, R. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, vol. 997. John Wiley & Sons, 2015.

[22] Kalivas, J. H. Two data sets of near infrared spectra. *Chemometrics and Intelligent Laboratory Systems 37*, 2 (1997), 255–259.

[23] Khademnoe, O., and Hosseini-Nasab, S. M. E. On asymptotic distribution of prediction in functional linear regression. *Statistics 50*, 5 (2016a), 974–990.

[24] Khademnoe, O., and Hosseini-Nasab, S. M. E. On properties of percentile bootstrap confidence intervals for prediction in functional linear regression. *Journal of Statistical Planning and Inference 170* (2016b), 129–143.

[25] Kokoszka, P., and Reimherr, M. *Introduction to Functional Data Analysis.* CRC Press, 2017.

[26] Kreiss, J.-P., and Franke, J. Bootstrapping stationary autoregressive moving-average models. *Journal of Time Series Analysis 13*, 4 (1992), 297–317.

[27] Morris, J. S. Functional regression. *Annual Review of Statistics and Its Application 2* (2015), 321–359.

[28] Pan, L., and Politis, D. N. Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions. *Journal of Statistical Planning and Inference 177* (2016), 1–27.

[29] Ramsay, J., and Silverman, B. W. *Functional Data Analysis.* Springer Series in Statistics. Springer-Verlag New York, 2005.

[30] Ratcliffe, S. J., Leader, L. R., and Heller, G. Z. Functional data analysis with application to periodically stimulated foetal heart rate data. I: functional regression. *Statistics in Medicine 21*, 8 (Apr. 2002), 1103–1114.

[31] Schlenker, W., and Roberts, M. J. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences 106*, 37 (Sept. 2009), 15594–15598. Publisher: National Academy of Sciences Section: Social Sciences.

[32] Wang, J.-L., Chiou, J.-M., and Müller, H.-G. Functional data analysis. *Annual Review of Statistics and its Application 3* (2016), 257–295.

[33] Wu, C.-F. J. Jackknife, bootstrap and other resampling methods in regression analysis. *the Annals of Statistics 14*, 4 (1986), 1261–1295.

[34] Yeon, H., Dai, X., and Nordman, D. J. Supplement to "bootstrap inference in functional linear regression models with scalar response".

# CHAPTER 3. BOOTSTRAP INFERENCE IN FUNCTIONAL LINEAR REGRESSION MODELS WITH SCALAR RESPONSE UNDER HETEROSCEDASTICITY

Modified from a manuscript submitted to *the Annals of Statistics*

Hyemin Yeon[1], Xiongtao Dai[2], and Daniel John Nordman[1]

[1]Department of Statistics, Iowa State University

[2]Division of Biostatistics, University of California, Berkeley

## Abstract

Inference for functional linear models in the presence of heteroscedastic errors has received insufficient attention given its practical importance; in fact, even a central limit theorem has not been studied in this case. At issue, conditional mean (projection of the slope function) estimates have complicated sampling distributions due to the infinite dimensional regressors, which create truncation bias and scaling problems that are compounded by non-constant variance under heteroscedasticity. As a foundation for distributional inference, we establish a central limit theorem for the estimated projection under general dependent errors, and subsequently we develop a paired bootstrap method to approximate sampling distributions. The proposed paired bootstrap does not follow the standard bootstrap algorithm for finite dimensional regressors, as this version fails outside of a narrow window for implementation with functional regressors. The reason owes to a bias with functional regressors in a naive bootstrap construction. Our bootstrap proposal incorporates debiasing and thereby attains much broader validity and flexibility with truncation parameters for inference under heteroscedasticity; even when the naive approach may be valid, the proposed bootstrap method performs better numerically. The bootstrap is applied to construct confidence intervals for projections and for conducting hypothesis tests for the slope

function. Our theoretical results on bootstrap consistency are demonstrated through simulation studies and also illustrated with real data examples.

## 3.1 Introduction

In classical linear models, bootstrap methods have been developed for several decades under either homoscedastic or heteroscedastic error assumptions. [22] first studied residual and paired bootstrap methods for approximating the sampling distribution of the least square estimator in multiple linear regression models. These bootstraps are intended, respectively, for handling homoscedastic or heteroscedastic error cases. Both bootstrap methods have been investigated in other contexts as well, for example, in nonparametric [29] or high-dimensional [20] regression problems.

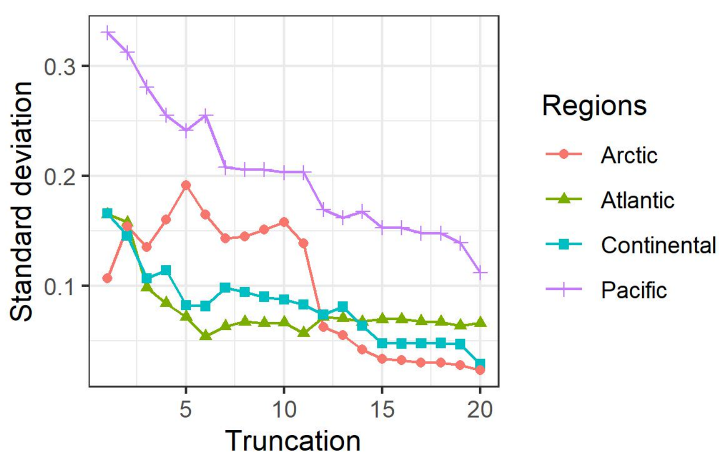


Figure 3.1: Estimated standard deviations for each region over different truncation levels used in estimation

In a functional linear regression model (FLRM), bootstrap inference is likewise valuable but also more complicated due to the infinite dimensionality of the underlying function space. A main issue with functional regressors is that a truncation bias arises in estimators of the conditional mean, because of the infinite dimensional regressor and slope function involved, which imposes challenges for even central limit theorems [12, 53]. Existing works on both the central limit theorem (CLT) and (residual/wild) bootstrap for functional linear regression models (FLRMs) have focused exclusively on homogeneous error variance models [25, 53], while either avoiding or accommodating this bias issue. In fact, beyond homoscedasticity assumptions, more stringent conditions of independence between regressors and errors are also commonly imposed in FLRM literature [6] and especially for hypothesis testing [8, 11, 31, 40, 41]. However, heteroscedastic error variances are commonly observed in practice. For illustration, Figure 3.1 shows the estimated standard deviations of residuals from an FLRM fit to a Canadian weather dataset (cf. Section 3.6) over different geographical regions. Each regressor curve represents averaged daily temperatures measured at a different location contained in one of the four regions in Canada: Atlantic, Continental, Pacific, and Arctic regions, where the associated response is the total annual precipitation on the log scale. As variances appear to differ across regions, it seems natural here to avoid homoscedastic error models.

To the best of our knowledge, heteroscedastic error conditions have not received much formal consideration in the FLRM literature, with perhaps the exception of work on weighted least squares by [19], which does not discuss distributional inference. For example, while a CLT for projection estimates is again available for FLRMs in the homoscedastic case [12, 34, 53], a counterpart foundational result does not yet exist under heteroscedasticity. One might further anticipate that previous bootstrap theory under homoscedasticity does not directly apply for the inference in FLRMs under heteroscedasticity. We show this to be the case, which necessitates our new development of a CLT and resampling theorems. As in the homoscedastic setting, resampling approximations in FLRMs are remarkably valuable under heteroscedasticity for capturing complicated sampling distributions of mean estimators, as current bootstraps from the homoscedastic case become invalid [25, 35, 53].

To bootstrap FLRMs in the presence of heteroscedastic errors, a paired bootstrap method can be considered, similar to the paired bootstrap for usual multiple linear regression models [22]. Paired bootstrap has indeed been applied for different inference in FLRMs [43, 46, 51], though without any theoretical development or justification. This latter point is important, because we show here that, surprisingly, a naive/standard implementation of paired bootstrap, adopted directly from the usual multiple regression case, can fail to provide valid inference for mean estimates under FLRMs if the truncation parameters are not set appropriately in a certain narrow and restricted way, in contrast to the case of finite-dimensional multiple linear regression [22]. In fact, as sample sizes increase with mean estimators in FLRMs, the distance between naive paired bootstrap and true sampling distributions may not converge to zero as typically expected, but rather can converge in distribution to a random number supported on $[0, 1]$ unless associated tuning parameters are set in a specific manner. The problem arises from a construction bias in the bootstrap world with FLRMs which relates to, but is a separate issue from, the truncation bias inherent to the FPCR estimator $\hat{\beta}_{h_n}$ of the slope. Such failure of bootstrap due to bias issues has been observed in other bootstrap works with complicated regressions, such as nonparametric [29, 30, 55], quantile [50], penalized linear [14, 15, 7], and high-dimensional linear [20] regression models, though the approaches of correcting bootstrap bias can differ. In some problems, the extent of the bias in paired bootstrap is such that this bootstrap must be discarded (cf. [29, 30]). This motivates our new development of paired bootstrap for FLRMs with heteroscedastic errors, which remedies the bias problem by modifying a bootstrap estimating equation to define a bootstrap estimator.

Under a general heteroscedastic error assumption, we study asymptotic and bootstrap inference in FLRMs with scalar response, along with providing its theoretical validity. In particular, we first establish a CLT for the projection estimator of $\langle \beta, X_0 \rangle$ with $X_0$ being a (random) new regressor function. This serves as the foundation for our bootstrap results and more broadly, justification of asymptotic inference for FLRMs under dependent errors. Our main bootstrap result is to develop a modified paired bootstrap to approximate the sampling distribution. We estimate the projection via the functional principal component regression (FPCR) estimator $\hat{\beta}_{h_n}$ of the slope function $\beta$ [9, 5, 28, 12, 25, 34, 35, 53], where $h_n$ denotes a truncation level involved in the estimation procedure. For flexibility and also for better practical performance, we allow additional truncation parameters $g_n, k_n$ to be introduced, and possibly vary from $h_n$, for defining important quantities in the bootstrap formulation, where $g_n$ arises to estimate bootstrap centering and $k_n$ is applied to estimate a scaling factor in studentization. The modified paired bootstrap incorporates an important debiasing step in order to accommodate a general combination of such truncation parameters. In the process of establishing a paired bootstrap, we also derive a new central limit theorem for the projection $\langle \beta, X_0 \rangle$ in FLRMs, involving an appropriate scaling for capturing different conditional error variances. In heteroscedastic cases, our numerical studies suggest that the paired bootstrap performs better than the residual bootstrap and normal approximation, while also maintaining good coverage in homoscedastic cases. The proposed paired bootstrap also numerically outperforms the naive version even when the latter is appropriately tuned. We consider intervals from studentization steps to obtain pivotal limits for use in bootstrap approximations. A rule of thumb for selecting the tuning parameters involved in the bootstrap procedure is further provided.

As an application of the paired bootstrap method, we treat a testing problem about the possible orthogonality of the slope function $\beta$ to subspaces spanned by a collection of target functional regressors. In this problem, the bootstrap combines several simultaneous estimation steps into one test, which would otherwise be distributionally intractable through normal approximations. The bootstrap construction also has the advantage of enforcing the null hypothesis in re-creating a reference distribution for testing, which can be useful for controlling size and boosting power. Our development in this testing problem is distinguishable from the previous works on hypothesis testing in FLRMs [8, 11, 31, 40, 41]: the latter tests are limited to independent error scenarios and often restrict claims to global nullity $\beta = 0$ or other specific projections of the slope function $\beta$ based on the cross-covariance between regressor and response. In contrast, by considering a general heteroscedastic setting, our work allows for formal hypothesis tests with FLRMs to be further justified under dependence between regressors and errors. Our testing method also allows hypotheses about $\beta$ be defined from projections with more arbitrarily specified functional regressors. This is useful in practice for assessing how projections of $\beta$ may differ from zero as predictor levels are varied, which may not be addressable by a global test of $\beta$, e.g., to examine effects of a hotter year on precipitation, or a higher winter/summer temperature contrast.

Section 3.2 describes the paired bootstrap method in FLRMs with scalar response under heteroscedasticity, along with a modification for general validity. Section 3.3 provides the main distributional results regarding estimated projections regarding the consistency of the paired bootstrap method and the failure of the naive bootstrap. With suitable scaling, a general CLT is also established, which is useful for framing studentized versions of statistics. We then give a consistent bootstrap procedure in Section 3.4 for testing the orthogonality over the slope function to linear subspaces. Numerical results are provided in Section 3.5, while Section 3.6 illustrates the paired bootstrap method with a real dataset that potentially have heteroscedasticity; an extra data application is provided in the supplement [54]. Some proofs for the main results are given in Appendix, while further details of the proofs and extended numerical results can be found in the supplement [54]. An R package is provided to construct confidence intervals for FLRM projections and to test the nullity of the projection of $\beta$ based on paired bootstrap.

## 3.2 Description of FLRMs and bootstrap

We start with the description of functional linear regression models (FLRMs) under heteroscedastic error variances in Section 3.2.1, and the paired bootstrap for estimated projections appears in Section 3.2.2.

### 3.2.1 FLRMs under heteroscedasticity

Consider the following FLRM

$$Y = \alpha + \langle \beta, X \rangle + \varepsilon, \tag{3.1}$$

where $Y$ is a scalar-valued response; $X$ is a regressor function taking values in a separable Hilbert space $\mathbb{H}$ with inner product $\langle \cdot, \cdot \rangle$; $\alpha$ is the intercept; and $\beta \in \mathbb{H}$ is the slope function. The error term $\varepsilon$ has $\mathsf{E}[\varepsilon|X] = 0$ but its distribution can otherwise depend on $X$; for example, heterogeneous conditional variances of the error $\varepsilon$ given the regressor $X$ is allowed, that is, $\sigma^2(X) \equiv \mathsf{E}[\varepsilon^2|X]$ may depend on the regressor $X$. As $\alpha = \mathsf{E}[Y] - \langle \beta, \mathsf{E}[X] \rangle$ holds in (3.1), without loss of generality, we assume that $\mathsf{E}[X] = 0$ and $\mathsf{E}[Y] = 0$ so that $\alpha = 0$ for purposes of developing estimation of the slope function $\beta$. The FLRM is then written as

$$Y = \langle \beta, X \rangle + \varepsilon. \tag{3.2}$$

Define the tensor product $x \otimes y : \mathbb{H} \times \mathbb{H} \to \mathbb{H}$ between two elements $x, y \in \mathbb{H}$ as a bounded linear operator $z \mapsto (x \otimes y)(z) = \langle z, x \rangle y$ for $z \in \mathbb{H}$. Under the assumption $\mathsf{E}[\|X\|^2] < \infty$ where $\| \cdot \|$ is the induced norm in $\mathbb{H}$, the covariance operator $\Gamma \equiv \mathsf{E}[(X \otimes X)]$ is self-adjoint, non-negative definite, and Hilbert–Schmidt, and hence, compact (cf. [33]). Then, $\Gamma$ admits the following spectral decomposition

$$\Gamma = \sum_{j=1}^{\infty} \gamma_j \pi_j$$

with $\pi_j \equiv \phi_j \otimes \phi_j$, where $\gamma_j$ and $\phi_j$ are the $j$-th eigenvalue and eigenfunction of $\Gamma$ for $j = 1, 2, \ldots$. Here, the set $\{\phi_j\}$ of eigenfunctions is an orthonormal system of $\mathbb{H}$ and $\{\gamma_j\}$ is a non-negative non-increasing sequence with $\gamma_j \to 0$ as $j \to \infty$. The functional version of normal equations is written as

$$\Delta = \Gamma \beta \tag{3.3}$$

from the model (3.2), where $\Delta \equiv \mathsf{E}[YX]$ is the cross-covariance function between $X$ and $Y$. Under the model identifiability assumption $\ker \Gamma = \{0\}$ [9, 10, 12] (see Assumption (A.1) of Section 3.3.1), the slope function is then given as

$$\beta = \Gamma^{-1} \Delta.$$

The functional principal component regression (FPCR) estimator of $\beta$ has been widely studied in the literature [5, 9, 12, 28]. To define the estimator, we suppose that the data pairs $\{(X_i, Y_i)\}_{i=1}^n$ are independently and identically distributed under the FLRM (3.2), that is,

$$Y_i = \langle \beta, X_i \rangle + \varepsilon_i, \quad i = 1, \ldots, n. \tag{3.4}$$

The sample versions of $\Gamma$ and $\Delta$ are defined as $\hat{\Gamma}_n \equiv n^{-1}\sum_{i=1}^n (X_i - \bar{X})^{\otimes 2}$ and $\hat{\Delta}_n \equiv n^{-1}\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})$, where $\bar{X} \equiv n^{-1}\sum_{i=1}^n X_i$, $\bar{Y} \equiv n^{-1}\sum_{i=1}^n Y_i$, and $x^{\otimes 2} \equiv x \otimes x$ for $x \in \mathbb{H}$. The sample covariance operator $\hat{\Gamma}_n$ also admits spectral decomposition $\hat{\Gamma}_n = \sum_{j=1}^n \hat{\gamma}_j \hat{\pi}_j$ with $\hat{\pi}_j \equiv \hat{\phi}_j \otimes \hat{\phi}_j$, where $\hat{\gamma}_j \geq 0$ is the $j$-th sample eigenvalue and $\hat{\phi}_j \in \mathbb{H}$ is the corresponding eigenfunction. By regularizing the inversion of $\hat{\Gamma}_n$, the FPCR estimator of $\beta$ is defined as

$$\hat{\beta}_{h_n} \equiv \hat{\Gamma}_{h_n}^{-1}\hat{\Delta}_n \tag{3.5}$$

where $\hat{\Gamma}_{h_n}^{-1} \equiv \sum_{j=1}^{h_n} \hat{\gamma}_j^{-1}\hat{\pi}_j$ is a finite approximation of $\Gamma^{-1} \equiv \sum_{j=1}^{\infty} \gamma_j^{-1}\pi_j$. Here, $h_n$ is the number of eigenpairs used in estimation, which represents a truncation level [5, 9, 12, 28].

### 3.2.2 Paired bootstrap procedure

For FLRMs with homoscedastic errors, the residual bootstrap is natural [25, 53], where this bootstrap re-creates data, e.g., $Y_i^* = \langle X_i, \hat{\beta}_{h_n}\rangle + \varepsilon_i^*$, through bootstrap error terms $\varepsilon^*$ as independent draws from an appropriate set of residuals. However, under heteroscedastic errors, a different bootstrap approach is necessary, akin to the standard multiple regression case with Euclidean vectors [22]. Similar to that setting for capturing response variances that may differ conditionally over regressors, we consider a paired bootstrap (PB) method for inference in FLRMs. To the best of our knowledge, the theory for PB in FLRMs has been studied only once by [24], but their application does not consider the slope function or its projections and the errors therein are homoscedastic in variance. For estimating means or projections under the FLRM with heteroscedastic errors, we explain next how the PB generally requires careful consideration in order to be valid.

To implement the PB, we draw the pairs $\{(X_i^*, Y_i^*)\}_{i=1}^n$ uniformly from the original data $\{(X_i, Y_i)\}_{i=1}^n$ with replacement. The bootstrap counterparts of sample moments are then given as $\hat{\Gamma}_n^* \equiv n^{-1}\sum_{i=1}^n (X_i^* - \bar{X}^*)^{\otimes 2}$ and $\hat{\Delta}_n^* \equiv n^{-1}\sum_{i=1}^n (Y_i^* - \bar{Y}^*)(X_i^* - \bar{X}^*)$ where $\bar{X}^* \equiv n^{-1}\sum_{i=1}^n X_i^*$ and $\bar{Y}^* \equiv n^{-1}\sum_{i=1}^n Y_i^*$. From the spectral decomposition of $\hat{\Gamma}_n^*$, we define a regularized inverse of

$\hat{\Gamma}_n^*$ with trunction level $h_n$ as

$$(\hat{\Gamma}_{h_n}^*)^{-1} \equiv \sum_{j=1}^{h_n} (\hat{\gamma}_j^*)^{-1} (\hat{\phi}_j^* \otimes \hat{\phi}_j^*),$$

where $\hat{\gamma}_j^*$ and $\hat{\phi}_j^*$ are the $j$-th eigenvalue and the corresponding eigenfunction of $\hat{\Gamma}_n^*$. This represents a direct bootstrap analog of $(\hat{\Gamma}_{h_n})^{-1}$ in (3.5).

An initial, though naive, bootstrap version $\hat{\beta}_{h_n,naive}^*$ of the FPCR estimator $\hat{\beta}_{h_n}$ can be found as

$$\hat{\beta}_{h_n,naive}^* \equiv (\hat{\Gamma}_{h_n}^*)^{-1} \hat{\Delta}_n^*$$

by directly imitating the definition of $\hat{\beta}_{h_n}$ in (3.5) with bootstrap data. The validity of this naive bootstrap, though, requires caution. The issue is that, in the bootstrap world, we need to define bootstrap version $\beta^*$ of the true parameter $\beta$ and, for flexibility, one might consider a FPCR estimator $\beta^* \equiv \hat{\beta}_{g_n} \equiv \hat{\Gamma}_{g_n}^{-1} \hat{\Delta}_n$ determined by a general truncation level $g_n$ in (3.5). It turns out that the naive bootstrap estimator $\hat{\beta}_{h_n,naive}^*$ must be restricted to a bootstrap parameter $\beta^* \equiv \hat{\beta}_{g_n}$ defined by $g_n = h_n$. The reason is due to a type of construction bias in the naive bootstrap, related to mimicking the linear structure in the model (3.2). Unless the bootstrap parameter $\beta^*$ is specifically chosen as $\hat{\beta}_{h_n}$, which imposes limitations for implementation and numerical performance, the naive bootstrap construction will be biased with a provably substantial and adverse effect on inference (cf. Proposition 13).

In order to define a more versatile bootstrap version $\hat{\beta}_{h_n}^*$ of the FPCR estimator $\hat{\beta}_{h_n}$, we begin from a general estimator, say $\hat{\beta}_{g_n}$, to play the role $\beta^*$ of the slope function $\beta$ in the bootstrap world; again $\hat{\beta}_{g_n}$ denotes an FPCR estimator similar to $\hat{\beta}_{h_n}$ from (3.5) but based on a truncation $g_n$ rather than $h_n$. The level of truncation $g_n$ used in a bootstrap version $\beta^* = \hat{\beta}_{g_n}$ of $\beta$ becomes a consideration because $\beta$ is infinite dimensional while any FPCR estimator $\hat{\beta}_{g_n}$ is finite-dimensional. It is possible to choose $g_n = h_n$, though more flexibility with $\beta^* = \hat{\beta}_{g_n}$ for $g_n$ smaller than $h_n$ can later provide a better re-creation of $\beta$ in the PB approximation than the original data estimator $\hat{\beta}_{h_n}$. However, regardless of the estimator $\hat{\beta}_{g_n}$ used to mimic $\beta$, the PB analog $\hat{\beta}_{h_n}^*$ of the original-data estimator $\hat{\beta}_{h_n}$ needs to be appropriately defined to avoid a construction bias in resampling. To correct this resampling bias, we adapt a modification of [48] for defining bootstrap M-estimators through adjusted bootstrap estimating equations; see also [37, 39] or [38], Section 4.3. It is non-trivial that this adjustment device should apply for PB in the FLRM case; in fact, this approach has not been applied in other bootstrap contexts where, for varying reasons, bootstrap construction has bias (e.g., nonparametric regression; [16]; lasso; [14, 15, 7]; high-dimensional regression; [20]).

To define a modified bootstrap version $\hat{\beta}^*_{h_n}$ of the FPCR estimator $\hat{\beta}_{h_n}$, we first observe that the slope function $\beta = \Gamma^{-1}\Delta$ can be prescribed as the solution to the estimating equation $\mathsf{E}[\Psi_i(\beta; \mu_X, \mu_Y)] = 0$ where

$$\Psi_i(\beta; \mu_X, \mu_Y) \equiv (X_i - \mu_X)(Y_i - \mu_Y) - (X_i - \mu_X)^{\otimes 2}\beta \tag{3.6}$$

is an estimating function with $\mu_X \equiv \mathsf{E}[X]$ and $\mu_Y \equiv \mathsf{E}[Y]$. A direct bootstrap counterpart of this estimating function is given by, say,

$$\check{\Psi}^*_i(\beta; \bar{X}, \bar{Y}) \equiv (X^*_i - \bar{X})(Y^*_i - \bar{Y}) - (X^*_i - \bar{X})^{\otimes 2}\beta,$$

where $\bar{X} \equiv n^{-1}\sum_{i=1}^n X_i$ and $\bar{Y} \equiv n^{-1}\sum_{i=1}^n Y_i$. A key observation is that, while $\beta = \Gamma^{-1}\Delta$ is the solution to the equation $\mathsf{E}[\Psi_i(\beta; \mu_X, \mu_Y)] = 0$, an estimator $\hat{\beta}_{g_n} \equiv \hat{\Gamma}^{-1}_{g_n}\hat{\Delta}_n$, playing the role of $\beta$ in the bootstrap world, will not generally be a solution to the equation

$$\hat{\Delta}_n - \hat{\Gamma}_n\beta \equiv \mathsf{E}^*[\check{\Psi}^*_i(\beta; \bar{X}, \bar{Y})] = 0$$

due to the finite dimensionality of $\hat{\beta}_{g_n}$, where $\mathsf{E}^*[\cdot] \equiv \mathsf{E}[\cdot|\mathcal{D}_n]$ denotes the bootstrap expectation conditional on the data $\mathcal{D}_n \equiv \{(X_i, Y_i)\}_{i=1}^n$. That is, due to truncation, $\hat{\Gamma}^{-1}_{g_n}$ does not generally match the inverse of $\hat{\Gamma}_n \equiv n^{-1}\sum_{i=1}^n (X_i - \bar{X})^{\otimes 2}$ for any finite truncation $g_n$. However, by starting from an estimator $\hat{\beta}_{g_n} \equiv \hat{\Gamma}^{-1}_{g_n}\hat{\Delta}_n$, we may adjust a bootstrap-level estimating function to be

$$\begin{aligned}\Psi^*_i(\beta; \bar{X}, \bar{Y}) &\equiv \check{\Psi}^*_i(\beta; \bar{X}, \bar{Y}) - \mathsf{E}^*[\check{\Psi}^*_i(\hat{\beta}_{g_n}; \bar{X}, \bar{Y})] \\ &= (X^*_i - \bar{X})(Y^*_i - \bar{Y}) - (X^*_i - \bar{X})^{\otimes 2}\beta - \hat{U}_{n,g_n},\end{aligned}$$

where

$$\mathsf{E}^*[\check{\Psi}_i^*(\hat{\beta}_{g_n};\bar{X},\bar{Y})] \equiv \hat{U}_{n,g_n} \equiv \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(\hat{\varepsilon}_{i,g_n}-\bar{\hat{\varepsilon}}_{g_n}) \tag{3.7}$$

has a closed form expression as the cross covariance between the regressors $\{X_i\}_{i=1}^n$ and the residuals $\{\hat{\varepsilon}_{i,g_n}\}_{i=1}^n$, $\hat{\varepsilon}_{i,g_n} \equiv Y_i - \langle \hat{\beta}_{g_n}, X_i\rangle$ arising from the estimator $\hat{\beta}_{g_n}$, with $\bar{\hat{\varepsilon}}_{g_n} \equiv n^{-1}\sum_{i=1}^n \hat{\varepsilon}_{i,g_n}$ above. These corrected bootstrap estimating functions have bootstrap expectation of

$$\mathsf{E}^*[\Psi_i^*(\beta;\bar{X},\bar{Y})] = \hat{\Delta}_n - \hat{\Gamma}_n\beta - \hat{U}_{n,g_n},$$

which equals zero at $\beta = \hat{\beta}_{g_n}$ in the bootstrap world. Consequently, $\hat{\beta}_{g_n}$ as the solution to $\mathsf{E}^*[\Psi_i^*(\beta;\bar{X},\bar{Y})] = 0$ mimics true slope function $\beta = \Gamma^{-1}\Delta$ solving $\mathsf{E}[\Psi_i(\beta;\mu_X,\mu_Y)] = 0$.

By replacing $\bar{X}$ and $\bar{Y}$ in $\Psi_i^*(\beta;\bar{X},\bar{Y})$ with bootstrap data counterparts $\bar{X}^* \equiv n^{-1}\sum_{i=1}^n X_i^*$ and $\bar{Y}^* \equiv n^{-1}\sum_{i=1}^n Y_i^*$ (in analog to the original estimator $\hat{\beta}_{h_n}$ defined by using $\bar{X}$ and $\bar{Y}$ in place of $\mu_X$ and $\mu_Y$), a PB version $\hat{\beta}^*_{h_n}$ of the FPCR estimator $\hat{\beta}_{h_n}$ is defined by the solution of the empirical bootstrap-data estimating equation

$$0 = \frac{1}{n}\sum_{i=1}^{n}\Psi_i^*(\beta;\bar{X}^*,\bar{Y}^*) = \hat{\Delta}_n^* - \hat{\Gamma}_n^*\beta - \hat{U}_{n,g_n},$$

upon regularization of $(\hat{\Gamma}_n^*)^{-1}$, where $\hat{\Gamma}_n^* \equiv n^{-1}\sum_{i=1}^n (X_i^*-\bar{X}^*)^{\otimes 2}$ and $\hat{\Delta}_n^* \equiv n^{-1}\sum_{i=1}^n (Y_i^*-\bar{Y}^*)(X_i^*-\bar{X}^*)$ are averages from the bootstrap sample. Hence, the PB re-creation of the FPCR estimator is then given by

$$\hat{\beta}^*_{h_n} \equiv (\hat{\Gamma}^*_{h_n})^{-1}(\hat{\Delta}_n^* - \hat{U}_{n,g_n}). \tag{3.8}$$

The construction in (3.8) matches how the original estimator $\hat{\beta}_{h_n} = \hat{\Gamma}_{h_n}^{-1}\hat{\Delta}_n$ from (3.5) is the solution of $0 = n^{-1}\sum_{i=1}^n \Psi_i(\beta;\bar{X},\bar{Y}) = \hat{\Delta}_n - \hat{\Gamma}_n\beta$, based on (3.6), upon similar regularization with truncation level $h_n$. The combination $(\hat{\beta}^*_{h_n}, \hat{\beta}_{g_n})$ in PB then serves to mimic $(\hat{\beta}_{h_n}, \beta)$ for inference about the FLRM.

## 3.3 Distributional results under heteroscedasticity

Section 3.3.1 first describes a CLT for estimated projections $\langle \hat{\beta}_{h_n}, X_0\rangle$ under the FLRM with heteroscedasticity. While novel and of potential interest in its own right, the CLT helps to develop the appropriate scaling needed for statistics and to also frame some baseline assumptions that are useful for bootstrap. Section 3.3.2 establishes the consistency of PB for distributional approximations. For comparison, Section 3.3.3 then provides a formal result to show that the naive implementation of bootstrap is generally invalid without restrictive conditions on truncation parameters.

### 3.3.1 CLT for the projections under heteroscedasticity

Let $X_0$ denote a new regressor under the model, which is independent of $\{(X_i, Y_i)\}_{i=1}^n$ and identically distributed as $X_1$. For an observed or given value of $X_0$ (i.e., conditional on $X_0$), we consider the sampling distribution of the difference

$$\sqrt{\frac{n}{s_{h_n}(X_0)}}[\langle \hat{\beta}_{h_n}, X_0\rangle - \langle \beta, X_0\rangle], \tag{3.9}$$

between estimated $\langle \hat{\beta}_{h_n}, X_0\rangle$ and true $\langle \beta, X_0\rangle$ projections. Above $s_{h_n}(X_0)$ denotes a scaling factor, based on $\Gamma_{h_n}^{-1} \equiv \sum_{j=1}^{h_n} \gamma_j^{-1}\pi_j$, which is defined as

$$s_{h_n}(x) \equiv \langle \Lambda\Gamma_{h_n}^{-1}x, \Gamma_{h_n}^{-1}x\rangle = \|\Lambda^{1/2}\Gamma_{h_n}^{-1}x\|^2, \quad x \in \mathbb{H}, \tag{3.10}$$

and involves the covariance operator $\Lambda \equiv \mathsf{E}[(X\varepsilon)^{\otimes 2}]$ of $X\varepsilon$, where $T^{1/2}$ denotes a self-adjoint square-root operator of a non-negative definite bounded linear operator $T$ on $\mathbb{H}$ such that $(T^{1/2})^2 = T^{1/2}T^{1/2} = T$. A sample counterpart of (3.10) is given as

$$\hat{s}_{h_n}(x) \equiv \langle \hat{\Lambda}_{n,k_n}\hat{\Gamma}_{h_n}^{-1}x, \hat{\Gamma}_{h_n}^{-1}x\rangle = \|\hat{\Lambda}_{n,k_n}^{1/2}\hat{\Gamma}_{h_n}^{-1}x\|^2, \quad x \in \mathbb{H}, \tag{3.11}$$

where $\hat{\Lambda}_{n,k_n} \equiv n^{-1}\sum_{i=1}^n \left(X_i\hat{\varepsilon}_{i,k_n} - n^{-1}\sum_{i=1}^n X_i\hat{\varepsilon}_{i,k_n}\right)^{\otimes 2}$ is an estimate of $\Lambda$ based on residuals $\hat{\varepsilon}_{i,k_n} \equiv Y_i - \langle \hat{\beta}_{k_n}, X_i\rangle$; for generality, here $k_n$ represents another tuning parameter used only to compute residuals $\{\hat{\varepsilon}_{i,k_n}\}_{i=1}^n$ for estimated scaling $\hat{s}_{h_n}(x)$ in (3.11).

Under either scaling factors $s_{h_n}(X_0)$ and $\hat{s}_{h_n}(X_0)$, we next show a CLT for the projection parameter $\langle \beta, X_0\rangle$ in Theorem 7, where the limiting distribution is standard normal under the scaling. For describing the CLT, some technical assumptions are listed.

(A1) $\ker\Gamma = \{0\}$, where $\ker\Gamma \equiv \{x \in \mathbb{H} : \Gamma x = 0\}$ ;

(A2) $\sup_{j\in\mathbb{N}} \gamma_j^{-2}\mathsf{E}[\langle X, \phi_j\rangle^4] < \infty$;

(A3) $\gamma_j$ is a convex function of $j \geq J$ (which implies that $\gamma_j - \gamma_{j+1}$ is decreasing) for some integer $J \geq 1$;

(A4) $\sup_{j\in\mathbb{N}} \gamma_j j \log j < \infty$;

(A5) $n^{-1}\sum_{j=1}^{h_n} \delta_j^{-2} \to 0$ as $n \to \infty$;

(A6) $h_n s_{h_n}(X)^{-1} = O_{\mathsf{P}}(1)$;

(A7) $\sup_{j\in\mathbb{N}} \lambda_j^{-2}\mathsf{E}[\langle X\varepsilon, \psi_j\rangle^4] < \infty$, where $\lambda_j$ and $\psi_j$ are the $j$-th eigenvalue–eigenfunction pair of $\Lambda$;

(A8) $\sup_{j\in\mathbb{N}} \gamma_j^{-1}\|\Lambda^{1/2}\phi_j\|^2 < \infty$.

Condition (A1) is necessary for the model identifiability [9, 10, 12]. Conditions (A2) and (A7) ensure that $X$ and $X\varepsilon$ respectively have finite fourth moments. Conditions (A3)-(A5) are technical assumptions related to the decay behaviors of eigenvalues $\{\gamma_j\}$ and eigengaps $\{\delta_j\}$, where for (A4) we define $\delta_1 \equiv \gamma_1 - \gamma_2$ and $\delta_j \equiv \min\{\gamma_j - \gamma_{j+1}, \gamma_{j-1} - \gamma_j\}$ for $j \geq 2$; such conditions are weak and are generally used to simplify proofs involving perturbation theory for functional data [12]. Condition (A6) provides a mild lower bound for scaling $s_{h_n}(X_0)$, where a similar assumption is needed in the homoscedastic setting; see [53] for a related discussion. Condition (A8) is a technical condition that balances the eignedecay of $\Gamma$ and the decay rate of $\Lambda$ in terms of $\{\phi_j\}_{j=1}^{\infty}$. When Condition (A2) holds, sufficient conditions for (A8) can also be developed by assuming moment structures on the error and regressors; for example,
Condition (A8) follows if either $\mathsf{E}[\varepsilon^4] < \infty$ or $\sigma^2(X) \equiv \mathsf{E}[\varepsilon^2|X] = \sum_{j=1}^{\infty}\rho_j^2\langle X, \phi_j\rangle^2$ for some $\{\rho_j\}_{j=1}^{\infty}$ such that $\sum_{j=1}^{\infty}\gamma_j\rho_j^2 < \infty$. The statement of the CLT also involves the following condition,
Condition $B(u)$ : $\sup_{j\in\mathbb{N}} j^{-1}m(j,u)\langle\beta, \phi_j\rangle^2 < \infty$,
depending on a generic constant $u > 0$ and function $m(j,u)$ of integer $j \geq 1$ defined as

$$m(j,u) = \max\left\{j^u, \sum_{l=1}^{j}\delta_l^{-2}\right\}. \tag{3.12}$$

Condition $B(u)$ is generally mild and helps to remove bias in the limiting distribution of the statistics from (3.9) by balancing the decay rates of eigenvalues and the Fourier coefficients of the slope function $\beta$, as described further in Remark 9 below.

A CLT for the projection in FLRMs under heteroscedasticity is a new development in the FLRM literature, as given in the following theorem.

**Theorem 7.** *Suppose that Conditions (A1)-(A7) hold along with* $h_n^{-1} + n^{-1/2} h_n^{7/2} (\log h_n)^3 \to 0$ *as* $n \to \infty$. *We further supppose* $n = O(m(h_n, u))$ *along with Condition* $B(u)$ *for some* $u > 7$. *Then, as* $n \to \infty$,

*(i)*

$$\sup_{y \in \mathbb{R}} \left| \mathsf{P}\left( \sqrt{\frac{n}{s_{h_n}(X_0)}} [\langle \hat{\beta}_{h_n}, X_0 \rangle - \langle \beta, X_0 \rangle] \leq y \middle| X_0 \right) - \Phi(y) \right| \xrightarrow{\mathsf{P}} 0,$$

*where* $\Phi$ *denotes the standard normal distribution function.*

*(ii) Additionally, if* $\|\hat{\beta}_{k_n} - \beta\| \xrightarrow{\mathsf{P}} 0$ *and Condition (A8) hold, then* $\hat{s}_{h_n}(X_0)$ *and* $s_{h_n}(X_0)$ *are asymptotically equivalent in that, for any* $\eta > 0$,

$$\mathsf{P}\left( \left| \frac{\hat{s}_{h_n}(X_0)}{s_{h_n}(X_0)} - 1 \right| > \eta \middle| X_0 \right) \xrightarrow{\mathsf{P}} 0,$$

*and the result in (i) also holds upon replacing* $s_{h_n}(X_0)$ *by the sample version* $\hat{s}_{h_n}(X_0)$.

Theorem 7 generalizes the CLT for projections in FLRMs [12, 53] from the homoscedastic case to broader heteroscedastic cases. When the errors are homoscedastic, i.e., $\mathsf{E}[\varepsilon^2|X] \equiv \sigma_\varepsilon^2 \in (0, \infty)$, then the covariance operator of $\epsilon X$ becomes $\Lambda = \sigma_\varepsilon^2 \Gamma$ and the scaling in (3.11) reduces to $s_{h_n}(X_0) = \sigma_\varepsilon^2 t_{h_n}(X_0)$, where $t_{h_n}(x) \equiv \|\Gamma_{h_n}^{-1/2} x\|^2$ for $x \in \mathbb{H}$. In this case, the result (i) in Theorem 7 matches the CLT under homoscedasticity [12, 53].

From Theorem 7, estimated projections $\langle \hat{\beta}_{h_n}, X_0 \rangle$ with data-based scaling $\hat{s}_{h_n}(X_0)$ are asymptotically pivotal and, hence, an asymptotic normal approximation may be applied to calibrate inference about $\langle \beta, X_0 \rangle$. However, resampling becomes useful for improving distributional approximations in FLRMs under heteroscedasticity, due to the complicated impacts of truncation $h_n$ in finite samples. The next section establishes the validity of the proposed PB method.

*Remark* 7. A sufficient condition for $\hat{\beta}_{k_n}$ to be consistent for $\beta$ in Theorem 7(ii) is that $k_n^{-1} + n^{-1/2}k_n^2 \log k_n \to 0$ as $n \to \infty$; see Theorem S1 of [54] for details. Theorem 7(ii) may be further generalized by replacing $\hat{\beta}_{k_n}$ used for constructing the estimated scaling $\hat{s}_{h_n}(X_0)$ with a general consistent estimator of $\beta$.

*Remark* 8. Under certain conditions on the error structure, the rate on the truncation level $h_n$ can be weakened to a lesser rate sufficient for obtaining a CLT under homoscedasticity. For instance, the rate $h_n^{-1} + n^{-1/2}h_n^{5/2}(\log h_n)^2 \to 0$ is sufficient for Theorem 7 if either $\mathsf{E}[\varepsilon^4] < \infty$ or $\mathsf{E}[\varepsilon^2|X] = \sum_{j=1}^{\infty} \rho_j^2 \langle X, \phi_j\rangle^2$ for some $\{\rho_j\}_{j=1}^{\infty}$ with $\sum_{j=1}^{\infty} \gamma_j \rho_j^2 < \infty$. This is the same rate as the ones assumed for the CLTs under homoscedasticity provided in [12, 53]. See Remark S1 in [54] for more details.

*Remark* 9. In Theorem 7 and Theorem 8 to follow, the Conditions $n = O(m(h_n, u))$ (or $n = O(m(g_n, u))$) and $B(u)$ are necessary only for removing bias in limit distribution of $\sqrt{n/s_{h_n}(X_0)}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ due to truncation $h_n$; that is, without these conditions, the asymptotic results would hold upon replacing $\langle\beta, X_0\rangle$ with a biased centering $\langle\Pi_{h_n}\beta, X_0\rangle$, where $\Pi_{h_n}$ denotes the projection on the first $h_n$ eigenfunctions $\{\phi_j\}_{j=1}^{h_n}$ of $\Gamma$ and $\Pi_{h_n}\beta \equiv \sum_{j=1}^{h_n}\langle\beta, \phi_j\rangle\phi_j$ is a truncated version of the slope $\beta \equiv \sum_{j=1}^{\infty}\langle\beta, \phi_j\rangle\phi_j$. Such conditions are common for balancing the decay rates of eigenvalues and the Fourier coefficients of the slope function in the removal of bias $\langle(\Pi_{h_n} - I)\beta, X_0\rangle$ from truncation. See [12, 53] for further discussion.

### 3.3.2 Consistency of the paired bootstrap (PB)

Based on scaling from the CLT for estimated projections in (3.13), we next consider PB approximations for the distribution of the studentized-type quantity

$$T_n(X_0) \equiv \sqrt{\frac{n}{\hat{s}_{h_n}(X_0)}}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle], \tag{3.13}$$

conditional on a given regressor $X_0$, involving estimated scaling $\hat{s}_n(X_0)$ from the data (3.11).

Based on the scaling factor $\hat{s}_n(X_0)$, one direct bootstrap counterpart of (3.13) is then given as

$$T^*_{n,\hat{s}}(X_0) \equiv \sqrt{\frac{n}{\hat{s}_{h_n}(X_0)}}[\langle \hat{\beta}^*_{h_n}, X_0\rangle - \langle \hat{\beta}_{g_n}, X_0\rangle] \tag{3.14}$$

with the same fixed $X_0$ fixed, where $(\hat{\beta}^*_{h_n}, \hat{\beta}_{g_n})$ denote the bootstrap analogs (3.8) of the FPCR estimator $\hat{\beta}_{h_n}$ and true slope $\beta$. Due to the shared scaling, this bootstrap version essentially approximates $[\langle \hat{\beta}_{h_n}, X_0\rangle - \langle \beta, X_0\rangle]$ with $[\langle \hat{\beta}^*_{h_n}, X_0\rangle - \langle \hat{\beta}_{g_n}, X_0\rangle]$.

To apply the bootstrap principle further, though, one might also consider a different bootstrap formulation of the studentized statistic (3.13) that aims to re-create the estimated scaling $\hat{s}_{h_n}(X_0)$ from bootstrap data. Recall that construction of $\hat{s}_{h_n}(\cdot)$ in (3.11) involves residuals from a FPCR estimator $\hat{\beta}_{k_n}$ with a generic bandwidth $k_n$. A bootstrap version of scaling factor is then defined, in analog to (3.11), as

$$\hat{s}^*_{h_n}(x) \equiv \langle \hat{\Lambda}^*_{n,k_n,g_n}(\hat{\Gamma}^*_{h_n})^{-1}x, (\hat{\Gamma}^*_{h_n})^{-1}x\rangle = \|(\hat{\Lambda}^*_{n,k_n,g_n})^{1/2}(\hat{\Gamma}^*_{h_n})^{-1}x\|^2, \ x \in \mathbb{H}, \tag{3.15}$$

where $\hat{\Lambda}^*_{n,k_n,g_n} \equiv n^{-1}\sum_{i=1}^n \left(X_i^*\hat{\varepsilon}^*_{i,k_n} - n^{-1}\sum_{i=1}^n X_i^*\hat{\varepsilon}^*_{i,k_n}\right)^{\otimes 2}$ is a bootstrap estimator of the covariance $\Lambda$ based on bootstrap residuals $\hat{\varepsilon}^*_{i,k_n} \equiv Y_i^* - \langle \hat{\beta}^*_{k_n}, X_i^*\rangle$ from a bootstrap FPCR estimator $\hat{\beta}^*_{k_n} \equiv (\hat{\Gamma}^*_{k_n})^{-1}(\hat{\Delta}^*_n - \hat{U}_{n,g_n})$; the latter is akin to (3.8) with tuning parameter $k_n$. A studentized bootstrap counterpart of (3.13), with estimated bootstrap scaling $\hat{s}^*_{h_n}(X_0)$, is then given as

$$T^*_{n,\hat{s}^*}(X_0) \equiv \sqrt{\frac{n}{\hat{s}^*_{h_n}(X_0)}}[\langle \hat{\beta}^*_{h_n}, X_0\rangle - \langle \hat{\beta}_{g_n}, X_0\rangle]. \tag{3.16}$$

Theorem 8 establishes the consistency of the PB method for the sampling distribution of the studentized projection estimator in (3.13) under heteroscedasticity. Let $\mathsf{P}^* \equiv \mathsf{P}(\cdot|\mathcal{D}_n)$ denotes the bootstrap probability conditional on the sample $\mathcal{D}_n \equiv \{(X_i, Y_i)\}_{i=1}^n$.

**Theorem 8.** *Suppose that Conditions (A1)-(A8) hold and that* $n^{-1}\sum_{j=1}^{h_n}\lambda_j^{-1} = O(1)$ *and* $k_n^{-1} + n^{-1/2}k_n^2 \log k_n \to 0$ *as* $n \to \infty$*. Along with Condition B(u) for some* $u > 7$*, we further suppose that* $\tau \equiv \lim_{n\to\infty} h_n/g_n \geq 1$, $g_n^{-1} + n^{-1/2}h_n^{7/2}(\log h_n)^3 \to 0$*, and* $n = O(m(h_n, u))$*. Then, as* $n \to \infty$*, the paired bootstrap (PB) is valid for the distribution of the studentized projection* *estimator* $T_n \equiv \sqrt{n/\hat{s}_{h_n}(X_0)}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$*:*

$$\sup_{y\in\mathbb{R}} \left|\mathsf{P}^*\left(T_n^*(X_0) \le y|X_0\right) - \mathsf{P}\left(T_n(X_0) \le y|X_0\right)\right| \xrightarrow{\mathsf{P}} 0,$$

*where* $T_n^*(X_0)$ *denotes either* $\hat{T}^*_{n,\hat{s}}(X_0)$ *from (3.14) or* $T^*_{n,\hat{s}^*}(X_0)$ *from (3.16).*

Theorem 8 conditions for the PB are similar to those for the CLT itself from Theorem 7, though additional mild assumptions (i.e., $\tau \equiv \lim_{n\to\infty} h_n/g_n \ge 1$) appear to govern the second truncation $g_n$ used in PB in conjunction with the original data truncation $h_n$. Namely, the truncation level $g_n$, for defining the bootstrap rendition $\hat{\beta}_{g_n}$ of the true parameter $\beta$, may differ from the other truncation level $h_n$ for defining the original FPCR estimator $\hat{\beta}_{h_n}$, though $g_n$ may not be larger than $h_n$ asymptotically (see also Proposition 14). This coordination of truncation levels is generally required for the bootstrap to be asymptotically correct, which allows the bootstrap to control the bias type described in Remark 9. In practice, we recommend choosing a slightly smaller $g_n$ than $h_n$. In particular, we give a rule of thumb for selecting $h_n$ and $g_n$ in Section 3.5, which performs well as illustrated numerically.

### 3.3.3 Limitations of naive bootstrap

As described in Section 3.2.2, a naive bootstrap formulation $\hat{\beta}^*_{h_n,naive} \equiv (\hat{\Gamma}^*_{h_n})^{-1}\hat{\Delta}^*_n$ of the FPCR estimator will not be generally be valid for approximating the distribution a projection estimator $T_n \equiv \sqrt{n/\hat{s}_{h_n}(X_0)}[\langle\hat{\beta}_{h_n}, X_0\rangle - \langle\beta, X_0\rangle]$ in (3.13) *unless* bootstrap centering parameter $\beta^* \equiv \hat{\beta}_{g_n}$ is narrowly chosen. That is, unlike with the PB method of Section 3.3.2 which is consistent when $\tau \equiv \lim_{n\to\infty} h_n/g_n \ge 1$ holds, the naive bootstrap in contrast can fail if $\tau \equiv \lim_{n\to\infty} h_n/g_n > 1$. This aspect arises due to an extra construction bias created in the naive bootstrap definition of $\hat{\beta}^*_{h_n,naive}$, particularly under heteroscedasticity. As a formal illustration, Proposition 13 considers a bootstrap quantity

$$T^*_{n,naive}(X_0) \equiv \sqrt{\frac{n}{\hat{s}_{h_n}(X_0)}}[\langle\hat{\beta}^*_{h_n,naive}, X_0\rangle - \langle\hat{\beta}_{g_n}, X_0\rangle] \tag{3.17}$$

that differs from a valid bootstrap version $T^*_{n,\hat{s}}(X_0)$ in (3.14) by using the naive bootstrap estimator $\hat{\beta}^*_{h_n,naive}$ in place of the proposed $\hat{\beta}^*_{h_n}$. In doing so, $T^*_{n,naive}(X_0)$ cannot capture the distribution of $T_n(X_0)$ if the heteroscedasticity is strong enough. Proposition 13 provides a general illustrative data example where the naive bootstrap method provably fails, which stands in contrast to the consistency of the modified PB from Theorem 8.

In the following, let $\mathbb{D}$ denote the space of all real-valued functions on $[-\infty, \infty]$ that are right continuous with left limits, which we equip with the Skorokhod metric (cf. [4]).

**Proposition 13.** *Suppose Theorem 8 assumptions along with* $n^{-1/2}h_n^{9/2}(\log h_n)^6 = o(1)$ *and* $\tau \equiv \lim_{n\to\infty} h_n/g_n > 1$. *We further suppose conditions (a)-(b) as follows:*

*(a) the conditional variance of the error* $\varepsilon$ *given the regressor* $X$ *is* $\sigma^2(X) \equiv \mathsf{E}[\varepsilon^2|X] = \sum_{j=1}^{\infty} \rho_j^2 \langle X, \phi_j\rangle^2$ *for some* $\{\rho_j\}_{j=1}^{\infty}$ *with* $\sum_{j=1}^{\infty} \gamma_j \rho_j^2 < \infty$;

*(b)* $X$ *has functional principal component (FPC) scores as* $\gamma_j^{-1/2}\langle X, \phi_j\rangle = \xi W_j$ *for* $j \geq 1$, *where* $\{W_j\}$ *denote iid standard normal variables and, independently,* $\xi$ *is a random variable with finite eighth moment* $\mathsf{E}[\xi^8] < \infty$;

*Then, the naive bootstrap version* $T^*_{n,naive}(X_0)$ *of* $T_n \equiv \sqrt{n/\hat{s}_{h_n}(X_0)}[\langle \hat{\beta}_{h_n}, X_0\rangle - \langle \beta, X_0\rangle]$ *satisfies that, as* $n \to \infty$,

$$\mathsf{P}^*(T^*_{n,naive}(X_0) \leq y|X_0) - \mathsf{P}(T_n(X_0) \leq y|X_0) \xrightarrow{d} \Phi\Big(y + \sigma(\tau)Z\Big) - \Phi(y), \quad y \in \mathbb{R},$$

*as elements in* $\mathbb{D}$, *where* $Z$ *denotes a standard normal variable with distribution function* $\Phi$ *and* $\sigma(\tau) > 0$ *denotes a constant (cf. (3.18)). Thus, the naive bootstrap is inconsistent.*

*Remark* 10. A take-away from Proposition 13 is that the naive bootstrap can fail with simple regressor structures, such as Gaussian $X$ (i.e., $\xi = 1$ above), though Condition (b) of Proposition 13 serves to accommodate a larger class of regressor distributions with potential dependence among FPCs.

*Remark* 11. The naive bootstrap (3.17) in Proposition 13 can be shown to be valid upon restricting $\tau \equiv \lim_{n\to\infty} h_n/g_n = 1$, which in case $\sigma(\tau) = 0$ (cf. (3.18)) so that the distributional limit becomes zero in the result. Essentially, bootstrap centering $\beta^* \equiv \hat{\beta}_{g_n}$ must be confined to the original FRCR estimator $\hat{\beta}_{h_n}$ (i.e., $g_n = h_n$). Further, neither the proposed or naive PB approach is generally valid if $\tau \equiv \lim_{n\to\infty} h_n/g_n < 1$ (cf. Proposition 14).

By Proposition 13, for increasing $n$, the distance between the true distribution function of the studentized projection estimator $T_n(X_0)$ and that of the naive bootstrap approximation $T^*_{n,naive}(X_0)$ does not converge to zero at any point on the real line, but rather behaves as a randomly drawn number in $(-1, 0)$ or $(0, 1)$ at each $y \in \mathbb{R}$. A similar bias issue, though, does not arise with standard applications of PB to regular finite-dimensional linear regression models (cf. [22]). A way to envision the bias of the naive bootstrap in FLRMs is as follows. From the proof of Proposition 13 and due to a construction bias, quantiles from the naive bootstrap approximation $T^*_{n,naive}(X_0)$ in (3.17) are shifted from those of a valid bootstrap approximation $T^*_{n,\hat{s}}(X_0)$ in (3.14) by a random contribution, say $B_n$, that depends on the original data but not the bootstrap sample; in large samples, this bias amount $B_n \approx T^*_{n,naive}(X_0) - T^*_{n,\hat{s}}(X_0)$ acts like a draw from a normal distribution with mean 0 and variance

$$\sigma^2(\tau) \equiv (1 - \tau^{-1}) \left( \|\Gamma^{1/2}\beta\|^2 \Big/ \sum_{j=1}^{\infty} \gamma_j \rho_j^2 + 1 \right) \tag{3.18}$$

where $\tau \equiv \lim_{n\to\infty} h_n/g_n \geq 1$ and so is non-ignorable if $\tau > 1$. See the supplement [54] for more details. This bias behavior can also be observed practically. Figure 3.2 contains a numerical illustration based on 1000 experiments generated from an FLRM with regressor $X$ and error $\varepsilon$ as described in Proposition 13. We examine the resulting distribution of the construction bias $B_n$ in the naive approach when $h_n/g_n > 1$. Figure 3.2 shows the distribution of this term $B_n$ is remarkably different from zero in small samples, even when $h_n = g_n + 1$. The bias $B_n$ is non-ignorable and becomes quite influential as the ratio $h_n/g_n$ becomes larger. The latter observation matches the theoretical result in (3.18), underling Proposition 13, in that the distributional spread of bias $B_n$ is greater as the ratio $h_n/g_n$ increases. Further simulation results in Section 3.5.1 indicate that naive bootstrap intervals also tend to over-cover.

For clarity, both naive and modified PB may fail if $\tau \equiv \lim_{n\to\infty} h_n/g_n < 1$ due to a different source of bias (i.e., apart from the construction of the bootstrap estimator $\hat{\beta}^*_{h_n}$), which relates to centering in the CLT (cf. Remark 9). This bias does not vanish if $h_n/g_n < 1$, which arises because

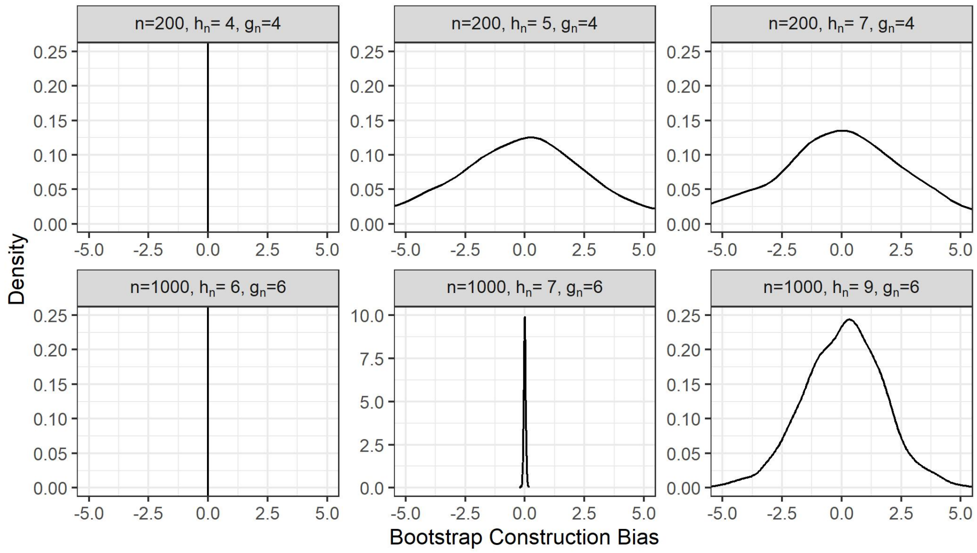


Figure 3.2: Kernel density estimates of a construction bias $B_n$ in the naive bootstrap. Plots have a common $x$-axis, and the bias is zero when $g_n = h_n$.

any estimator $\hat{\beta}_{g_n}$, playing the role of the true slope $\beta$ in the bootstrap world, cannot capture the infinite dimensionality of $\beta$. This failure is illustrated in Proposition 14, with details in the supplement [54].

**Proposition 14.** *Suppose the assumptions of Theorem 8 along with* $\tau \equiv \lim_{n\to\infty} h_n/g_n \in (0,1)$. *We further suppose Conditions (a) and (b) in Proposition 13. Then, as* $n \to \infty$, *both naive* $T^*_{n,naive}(X_0)$ *and modified* $T^*_n(X_0)$ *bootstrap renditions of* $T_n \equiv \sqrt{n/\hat{s}_{h_n}(X_0)}[\langle \hat{\beta}_{h_n}, X_0\rangle - \langle \beta, X_0\rangle]$ *fail to provide asymptotically correct distributional approximations. Namely, the convergence in Proposition 13 holds for* $T^*_{n,naive}(X_0)$ *upon redefining* $\sigma(\tau)$ *there as* $\sqrt{\tau^{-1} - 1} > 0$, *and this same result holds also for* $T^*_n(X_0)$.

To summarize our bootstrap distributional results, the asymptotic ratio $\tau \equiv \lim_{n\to\infty} h_n/g_n \in (0,\infty)$ plays a significant role in PB methods. If $h_n$ is asymptotically larger than $g_n$ with $\tau > 1$, our modified PB is applicable (Theorem 8) while the naive PB may provably fail (Proposition 13), even with simple FPC scores. If $\tau = 1$, both PB methods work, though the naive one requires a stronger tuning parameter condition $n^{-1/2}h_n^{9/2}(\log h_n)^6 = O(1)$ (cf. Remark 11). Thus, the naive PB requires asymptotic equivalence of $h_n$ and $g_n$ and becomes invalid when $h_n$ is asymptotically larger than $g_n$. In contrast, the modified PB enjoys additional flexibility in setting $h_n$ and $g_n$, which will be shown to result in better numerical performance in Section 3.5. Namely, simulation evidence indicates that our modified PB produces good and stable coverages over different $h_n \geq g_n$, while the performance of the naive PB can be more sensitive to the ratio $h_n/g_n$. Finally, if $h_n$ is asymptotically smaller than $g_n$ with $\tau < 1$, both PB methods might fail due to the bias from the finite dimensionality of the estimator $\hat{\beta}_{g_n}$, serving as the bootstrap rendition of the true slope parameter $\beta$ (Proposition 14).

*Remark* 12. The bootstrap theory in Theorem 8 and Proposition 13 can be generalized in that the ratio $h_n/g_n$ need not have a convergent limit, and $\tau \equiv \lim_{n\to\infty} h_n/g_n$ may be replaced by $\tau \equiv \liminf_{n\to\infty} h_n/g_n$.

### 3.4 Hypothesis tests for projections onto finite-dimensional subspaces

The testing of the association between the functional regressor $X$ and the scalar response $Y$ in FLRMs has drawn much recent attention in the literature. [8] first proposed a global test of $\beta = 0$ by assessing the covariance operator for $\Delta = 0$, and several works have similarly considered various global tests [11, 24, 31, 40, 41]. In contrast, [36] and [49] focused on testing of $\langle \beta, \phi_1 \rangle = \cdots = \langle \beta, \phi_L \rangle = 0$ for some pre-fixed integer $L$, where $\{\phi_j\}_{j=1}^{\infty}$ denotes the set of eigenfunctions of $\Gamma \equiv \mathsf{E}[(X - \mathsf{E}[X])^{\otimes 2}]$. However, none of these previous works applies to testing the orthogonality of $\beta$ to generally specified regressor points. In this testing problem with FLRMs, though, the proposed PB method from Section 3.2 can be adapted to assess projections of the slope function $\beta$ onto subspaces spanned by general directions.

To frame the testing problem, let $\mathcal{X}_0 \equiv \{X_{0,l}\}_{l=1}^{L}$ denote a collection of regressors $X_{0,1}, \ldots, X_{0,L}$ under consideration, for some integer $L \geq 1$. Letting $\Pi_{\mathcal{X}_0}\beta$ denote the projection of the slope function $\beta$ onto the linear subspace $\mathrm{span}(\mathcal{X}_0) \subseteq \mathbb{H}$ spanned by $\mathcal{X}_0$, we wish to test the null hypothesis

$$H_0 : \Pi_{\mathcal{X}_0}\beta = 0 \quad \text{against} \quad H_1 : H_0 \text{ is not true},$$

regarding the orthogonality of $\beta$ to span$(\mathcal{X}_0)$. The null hypothesis is equivalently stated as $\langle \beta, X_{0,l}\rangle = 0$ for all $l = 1, \ldots, L$. As the bootstrap results in Section 3.3 apply for a given regressor $X_0$, a PB-based testing procedure can be formulated to assess this type of hypothesis. An advantage is that this approach provides a specific test of whether regression effects exist in any pre-defined directions, while a global test about $\beta$ (e.g., based on the covariance $\Delta$) is not amenable to this purpose. Additionally, essentially all of the previous works on hypothesis testing for FLRMs rely on independent error assumptions, while our bootstrap-based testing procedure provides a first work on addressing such testing problems in FLRMs under dependent errors and heteroscedasticity.

To describe test statistics, write

$$T_{n,l}^{H_0} \equiv \sqrt{\frac{n}{\hat{s}_{h_n}(X_{0,l})}} \langle \hat{\beta}_{h_n}, X_{0,l}\rangle, \quad l = 1, \ldots, L, \tag{3.19}$$

to denote the studentized projection estimator (3.13) for each direction with centering $\langle X_{0,l}, \beta\rangle = 0$ under the null hypothesis. We may define test statistics by combining these direction-based statistics as

$$W_{n,L^2} \equiv \sum_{l=1}^{L} \left[T_{n,l}^{H_0}\right]^2 \quad \text{and} \quad W_{n,\max} \equiv \max_{1\le l\le L} \left|T_{n,l}^{H_0}\right|, \tag{3.20}$$

representing $L_2$- or $L_\infty$-type norms. Large values of such statistics then provide evidence against the null hypothesis. While both test statistics are well-defined with non-degenerate limit distributions under the null hypothesis, these limit laws are complicated under heteroscedasticity, depending intricately on covariances between estimated projections. Consequently, these limit distributions are impractical for direct use. However, the sampling distributions of test statistics can be viably approximated with the proposed PB method and, in fact, there exist two ways of implementing the bootstrap here: by enforcing the null hypothesis at the bootstrap level or not.

If we do not enforce the null hypothesis in the bootstrap world, then we essentially adopt the same PB procedure described in Section 3.3.2 (i.e., Theorem 8). That is, we may formulate studentized bootstrap quantities, similar to (3.16), as

$$T_{n,l,\hat{s}^*}^{*} \equiv \sqrt{\frac{n}{\hat{s}_{h_n}^*(X_{0,l})}} [\langle \hat{\beta}_{h_n}^*, X_{0,l}\rangle - \langle \hat{\beta}_{g_n}, X_{0,l}\rangle], \quad l = 1, \ldots, L$$

based on the same bootstrap sample $\{(X_i^*, Y_i^*)\}_{i=1}^n$ and a common estimator $\hat{\beta}_{g_n}$ playing the bootstrap role of $\beta$. The bootstrap test statistics are then given by

$$W_{n,L^2}^* \equiv \sum_{l=1}^{L} \left[T_{n,l,\hat{s}^*}^*\right]^2 \quad \text{and} \quad W_{n,\max}^* \equiv \max_{1\leq l\leq L} \left|T_{n,l,\hat{s}^*}^*\right|, \tag{3.21}$$

To enforce the null hypothesis in the bootstrap world, we modify the PB procedure described in Section 3.2.2, letting $\tilde{\beta}_{g_n} \equiv \hat{\beta}_{g_n} - \Pi_{\mathcal{X}_0}\hat{\beta}_{g_n}$ rather than $\hat{\beta}_{g_n}$ denote the bootstrap analog of the slope $\beta$. Here $\tilde{\beta}_{g_n}$ denotes a version of $\hat{\beta}_{g_n}$ after removing its projection $\Pi_{\mathcal{X}_0}\hat{\beta}_{g_n}$ onto the subspace spanned by $\mathcal{X}_0$. With this change, it holds that $\Pi_{\mathcal{X}_0}\tilde{\beta}_{g_n} = 0$ and so $\tilde{\beta}_{g_n}$ mimics the same property $\Pi_{\mathcal{X}_0}\beta = 0$ of the true parameter $\beta$ under $H_0$. To formulate bootstrap data, we also write a response variable $\tilde{Y}_i \equiv Y_i - \langle \Pi_{\mathcal{X}_0}\hat{\beta}_{g_n}, X_i\rangle$ after removing a projection contribution from $\Pi_{\mathcal{X}_0}\hat{\beta}_{g_n}$ with respect to $X_i$. A PB sample $\{(X_i^*, \tilde{Y}_i^*)\}_{i=1}^n$ is defined by iid draws from the empirical distribution of $\{(X_i, \tilde{Y}_i)\}_{i=1}^n$, and the same development from Section 3.2.2 then applies with the change that $Y_i^*$, $\hat{\beta}_{g_n}$, $\bar{Y}$ there become $\tilde{Y}_i^*$, $\tilde{\beta}_{g_n}$, $\bar{\tilde{Y}} \equiv n^{-1}\sum_{i=1}^n \tilde{Y}_i = \bar{Y} - \langle \Pi_{\mathcal{X}_0}\hat{\beta}_{g_n}, \bar{X}\rangle$. That is, a baseline estimating function becomes

$$\check{\Psi}_i^*(\beta; \bar{X}, \bar{\tilde{Y}}) \equiv (X_i^* - \bar{X})(\tilde{Y}_i^* - \bar{\tilde{Y}}) - (\tilde{X}_i^* - \bar{X})^{\otimes 2}\beta,$$

and its bootstrap expectation at $\tilde{\beta}_{g_n}$ is $\mathsf{E}^*[\check{\Psi}_i^*(\tilde{\beta}_{g_n}; \bar{X}, \bar{\tilde{Y}})] \equiv \hat{U}_{n,g_n}$, similar to (3.7); a mean-corrected estimating function is then $\Psi_i^*(\beta; \bar{X}, \bar{\tilde{Y}}) \equiv \check{\Psi}_i^*(\beta; \bar{X}, \bar{\tilde{Y}}) - \hat{U}_{n,g_n}$; and the bootstrap version $\tilde{\beta}_{h_n}^*$ of the original FPCR estimator $\hat{\beta}_n$ is the (regularized) solution to the bootstrap empirical average $n^{-1}\sum_{i=1}^n \check{\Psi}_i^*(\beta; \bar{X}^*, \bar{\tilde{Y}}^*) = 0$ with $\bar{X}^* \equiv n^{-1}\sum_{i=1}^n X_i^*$ and $\bar{\tilde{Y}}^* \equiv n^{-1}\sum_{i=1}^n \tilde{Y}_i^*$ from the bootstrap sample. The bootstrap estimator then has a closed form as

$$\tilde{\beta}_{h_n}^* \equiv (\hat{\Gamma}_{h_n}^*)^{-1}(\tilde{\Delta}_n^* - \hat{U}_{n,g_n})$$

in parallel to (3.8) with $\tilde{\Delta}_n^* \equiv n^{-1}\sum_{i=1}^n (\tilde{Y}_i^* - \bar{\tilde{Y}}^*)(X_i^* - \bar{X}^*)$ in place of $\hat{\Delta}_n^* \equiv n^{-1}\sum_{i=1}^n (Y_i^* - \bar{Y}^*)(X_i^* - \bar{X}^*)$. When enforcing the null hypothesis at the bootstrap level, bootstrap versions of test statistics in (3.20) are then given by

$$W_{n,L^2}^* \equiv \sum_{l=1}^{L} \left[T_{n,l}^{*H_0}\right]^2 \quad \text{and} \quad W_{n,\max}^* \equiv \max_{1\leq l\leq L} \left|T_{n,l}^{*H_0}\right|, \tag{3.22}$$

with

$$T_{n,l}^{*H_0} \equiv \sqrt{\frac{n}{\tilde{s}^*_{h_n}(X_{0,l})}} \langle \tilde{\beta}^*_{h_n}, X_{0,l} \rangle, \quad l = 1, \dots, L,$$

denoting the bootstrap rendition of the estimated projection quantities $T_{n,l}^{H_0}$ from (3.19) under $H_0$. Above $\tilde{s}^*_{h_n}$ denotes estimated scaling, akin to $\hat{s}_{h_n}$, computed from the bootstrap sample $\{(X_i^*, \tilde{Y}_i^*)\}_{i=1}^n$.

The following result guarantees that, under the null hypothesis $H_0 : \Pi_{\mathcal{X}_0}\beta = 0$, the distribution of test statistics $W_{n,L^2}$ and $W_{n,\max}$ in (3.20) can be approximated by either bootstrap approach: enforcing $H_0$ as in (3.22) or not as in (3.21).

**Corollary 7.** *Let $W_n$ denote a test statistic $W_{n,L^2}$ or $W_{n,\max}$ and let $W_n^*$ denote its paired bootstrap counterpart, computed either as in (3.21) or (3.22). Under the assumptions of Theorem 8, if the null hypothesis $H_0 : \Pi_{\mathcal{X}_0}\beta = 0$ holds, then*

$$\sup_{w \in \mathbb{R}} |\mathsf{P}^*(W_n^* \leq w | \mathcal{X}_0) - \mathsf{P}(W_n \leq w | \mathcal{X}_0)| \xrightarrow{\mathsf{P}} 0 \quad \text{as } n \to \infty.$$

While both implementations (3.21)-(3.22) of PB are valid for testing, numerical studies suggest that enforcing the null hypothesis (3.22) can have better performance in both size and power. This is explored further in Section 3.5.2.

## 3.5 Simulation studies

Section 3.5.1 summarizes numerical studies of the PB and other methods for calibrating confidence intervals for projections in FLRMs. A rule of thumb for selecting the tuning parameters $(k_n, h_n, g_n)$ in the bootstrap procedure is also examined. Section 3.5.2 then investigates the performance of the bootstrap test from Section 3.4 regarding projections.

### 3.5.1 Performance of bootstrap intervals

Here we examine, through simulation, PB confidence intervals for a projection $\langle \beta, X_0 \rangle$. To describe the data generation, we independently simulate $n$ curves $\mathcal{X}_n = \{X_i\}_{i=1}^n$ from a truncated

Karhunen–Loève expansion:

$$X \stackrel{\mathrm{d}}{=} \sum_{j=1}^{J} \sqrt{\gamma_j}\xi_j\phi_j \tag{3.23}$$

with $J = 15$. Above $\{\phi_j : j = 1, \ldots, J\}$ denote the first $J$ of the Fourier basis functions $\{1, \sin(2\pi t), \cos(2\pi t), \ldots\}$ on $[0, 1]$. The FPC scores are defined as $\xi_j = \xi W_j$, where $W_j \stackrel{\text{iid}}{\sim} \mathsf{N}(0, 1)$ and $\xi$ follows a $t(\nu)$ distribution with chosen degrees of freedom $\nu \in \{4, 5, 7, 9, \infty\}$. This entails that FPC scores are uncorrelated, but dependent. The eigengaps in (3.23) are defined with a polynomial decay rate involving a parameter $a > 0$, namely $\gamma_j - \gamma_{j+1} = 2j^{-a}$, $j \geq 1$ where $\gamma_1 = \sum_{j=1}^{\infty} j^{-a}$. Using the same basis functions, the slope parameter is set to $\beta = \sum_{j=1}^{J} \beta_j\phi_j$, where $\beta_j = 3j^{-b}W_{\beta,j}$ has a polynomial decay involving a rate parameter $b > 0$ and the terms $W_{\beta,j}$ are fixed upon drawing these as iid from a distribution $\mathsf{P}(W_{\beta,j} = 1) = 1/2 = \mathsf{P}(W_{\beta,j} = -1)$. We consider various scenarios involving different polynomial rates and sample sizes: $a, b \in \{1.5, 2.5, 3.5, 4.5, 5.5\}$ and $n \in \{50, 200, 1000\}$. For brevity, we report some representative numerical results here, though full results can be found in the supplement [54]. All the function values are evaluated at 100 equally-spaced time grid points in $[0, 1]$. Response values $\{Y_i\}_{i=1}^{n}$ are then generated through the FLRM (3.4) as follows. To consider both homoscedastic and heteroscedastic scenarios, errors $\varepsilon_i$ are generated to be either independent from or dependent on the regressors $X_i$. For a given generated regressor $X_i$, a dependent error $\varepsilon_i$ is simulated from a chi-square distribution $\chi^2(\nu(X_i)) - \nu(X_i)$ with $\nu(X_i) \equiv \|X_i\|^2/2$ degrees of freedom. In this heteroscedastic case, the conditional variance of an error depends on the regressor value $X_i$, and the marginal variance of an error is $\mathsf{var}[\varepsilon_i^2] = \mathrm{tr}\,(\Gamma) = \sum_{j=1}^{J} \gamma_j$. Due to the latter, we also generate errors $\varepsilon_i$ with the same marginal variance, independently from regressors $X_i$, with a centered chi-square distribution $\chi^2(\nu) - \nu$ with $\nu \equiv \mathrm{tr}\,(\Gamma)/2$ degrees of freedom in homoscedastic cases. The supplement [54] provides further results with other error distributions, which are qualitatively similar. In each simulation run, a regressor $X_0$ for projection estimation is also generated by (3.23).

We consider both PB and naive PB implementations for computing two-sided 95% intervals for a projection $\langle X_0, \beta\rangle$. In the original data FPCR estimator $\hat{\beta}_{h_n}$ from (3.5) and estimated scaling $\hat{s}_n$ from (3.11), we varied the range of the truncation parameters $h_n, g_n \in \{1, \ldots, 15\}$ and we set $k_n = 2[n^{1/v}]$ with $v = 2a + 1 + v_1$ for a small $v_1 = 0.1$ for consistent estimation of $\hat{\beta}_{k_n}$ in scaling (3.11) (cf. Theorem S1 of the supplement [54]). To recall, $k_n$ is used to reconstruct the residual as used in the scaling factor (3.11), $g_n$ is for constructing the bootstrap centering, and $h_n$ is the truncation used by the actual and bootstrap estimators (see Theorem 8). For simplicity here, we focus on symmetrized intervals in the PB implementation involving bootstrap studentization (e.g., $T^*_{n,\hat{s}^*}(X_0)$ in (3.16)) as well as a naive PB counterpart defined by replacing scaling $\hat{s}_{h_n}(X_0)$ in (3.17) with a bootstrap sample counterpart $\hat{s}^*_{h_n}(X_0)$ from $\hat{\beta}^*_{k_n,naive} = (\hat{\Gamma}^*_{k_n})^{-1}\hat{\Delta}^*_n$; further comparisons with non-studentized versions of PB (e.g., $T^*_{n,\hat{s}}(X_0)$ in (3.14)) or non-symmetrized intervals can be found in the supplement [54], though bootstrap studentization steps tend to induce the best performances. For comparison, we also consider intervals based on normal approximations with estimated scaling $\hat{s}_n$ (Theorem 7) or residual bootstrap (RB) (cf. [53]). For each generated data set, bootstrap distributions are approximated by 1000 Monte Carlo resamples.

We also propose a rule of thumb for setting the tuning parameters based on simulations for all the parameter combinations. We suggest to set $g_n = k_n$ and $h_n = [1.113k_n]$ being a slightly larger value than $g_n$; the value of $k_n$ can be selected in practice by cross-validation minimizing the prediction errors. Our rule of thumb is found by considering all scenarios and truncation levels producing coverages of PB intervals within 1% from the nominal level 95%, and running linear regression of response $(h_n, g_n)$ on $k_n$. This rule targets to make appropriate choices of $(h_n, g_n)$, as most critical to performance of PB, in relation to $k_n$. Setting $g_n = k_n$ aligns with the appropriate choices for the RB [53].

For each 95% interval procedure for $\langle X_0, \beta\rangle$, empirical coverages were approximated by 1000 simulation runs over each data generating model and sample size. Figure 3.3 displays observed coverage rates from different methods under a few selected scenarios when $a = 2.5$, $b = 5.5$ and $\xi \sim t(5)$; see the supplement [54] for results over all scenarios. For clarity, the results in Figure 3.3

focus on the case that $g_n = k_n$ for both PB and RB while varying $h_n$. Coverages for the PB method under the proposed rule of thumb are indicated using crosses in Figure 3.3 for reference.

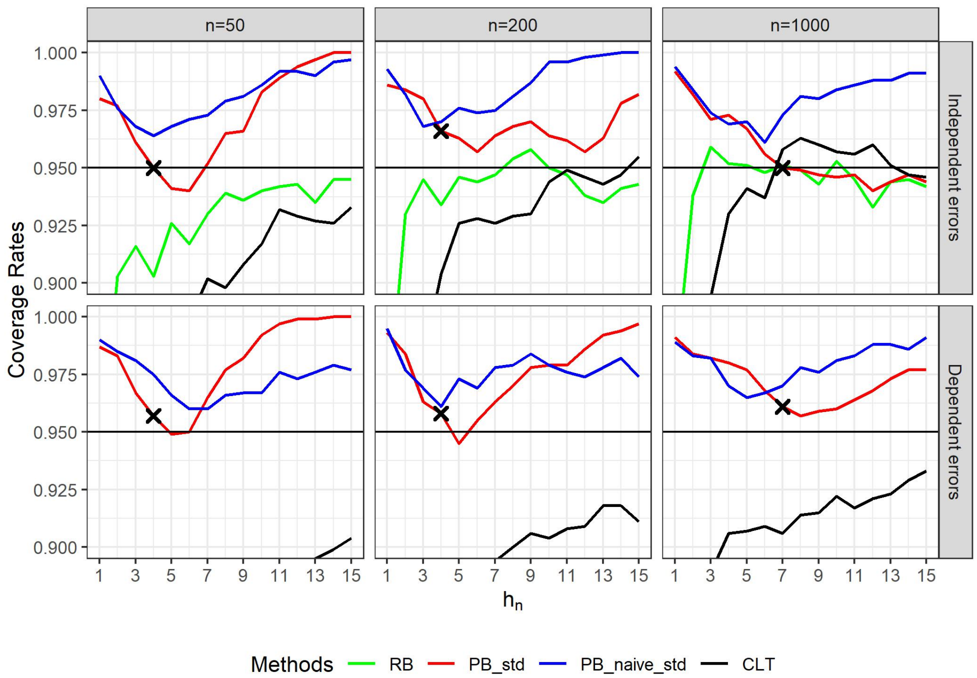


Figure 3.3: Empirical coverage rates of 95% intervals for $\langle \beta, X_0 \rangle$ from CLT (black), RB (green), PB with studentization (red), and naive PB with studentization (blue), over various truncations $h_n$ when the decay rates for $\gamma_j - \gamma_{j+1}$ and $\beta_j$ are set to be $a = 2.5$ and $b = 5.5$ and the latent variable for the FPC scores is $\xi \sim t(5)$. In particular, for errors dependent on regressors (lower panels), the coverage curves of CLT/RB intervals are cropped as these perform poorly. Crosses $\times$ indicate coverage rates with $h_n$ selected by the proposed rule.

As a first observation from Figure 3.3, the coverages from intervals based directly on normal approximation (CLT) exhibit sensitivity to the truncation level $h_n$ and also under coverage, particularly when the sample size is small. Under heteroscedasticity, both the CLT and RB methods perform quite poorly and lie at least partially outside of the charting regions in Figure 3.3. In fact, RB is not asymptotically valid in this case and the coverages are quite low to the extent that coverage curves do not appear in the figure, even for large sample sizes $n = 1000$. In contrast, PB intervals perform much better under the heteroscedastic models. For independent errors, while RB assumes and uses the true model structure (homoscedasticity) and PB does not, the PB method has very similar performance to RB for large sample sizes ($n = 1000$) and exhibits comparable performance for smaller samples ($n = 50$ or 200). Our rule of thumb provides reasonable coverages in most cases for PB intervals.

Figure 3.4 displays the corresponding average widths of intervals, which generally increase with $h_n$. Importantly, this figure indicates that intervals from RB and CLT approximations are often overly narrow under heteroscedasticity, which relates to the low coverages in Figure 3.3. Figures 3.3-3.4 also demonstrate that our rule of thumb seems to suggest an optimal truncation $h_n$ in the sense that the corresponding intervals balance good coverage rates with lowest average widths. Finally, while the naive PB implementation is asymptotically invalid in the sense of Proposition 13, the latter finding also suggests that the bias in the naive PB should translate to over-coverage for symmetrized intervals in Figure 3.3. Even for large sample sizes $n = 1000$, naive PB intervals tend to over-cover projections, while their average widths are larger than those from the proposed PB. Moreover, the coverages of naive PB intervals are unstable against the choice of truncation level $h_n$ while our modified PB produces stable coverages close to the nominal level for all $h_n \geq g_n$ and moderate to large sample sizes $n = 200$ and 1000. The over-coverage problem in the naive PB also worsens as truncation levels $h_n$ deviate from the case $h_n = g_n$. This can be interpreted as the construction bias from the naive bootstrap negatively impacts this method, even as the sample size increases.

*Remark* 13. To investigate the effect of the moments of the regressor $X$ on interval performance, we also varied the distribution of $\xi$ in (3.23) over different $t(\nu)$ cases with $\nu \in \{4, 5, 7, 9, \infty\}$, where $t(4)$ provides an example that does not satisfy (A2). Figures in Section S3.1 of the supplement [54] show that, under heteroscedasticity, both unsymmetrized and symmetrized intervals from the proposed PB (either with or without bootstrap studentization steps) are fairly robust to the moment of $X$, while the RB is quite sensitive to the number of finite moments of $X$; the coverages of the RB method tend to increase to the nominal level as more moments for $\xi$ become available, though the coverages remain quite low. However, regardless of the available

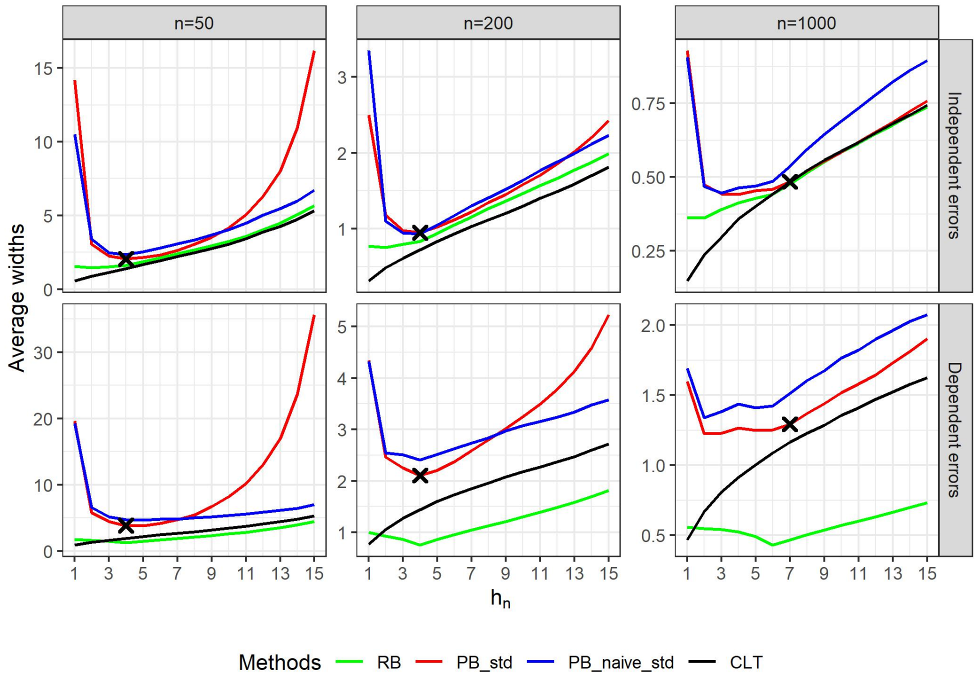


Figure 3.4: Average widths of 95% intervals for $\langle \beta, X_0 \rangle$ from methods over different truncation $h_n$ when the decay rates for $\gamma_j - \gamma_{j+1}$ and $\beta_j$ are set to be $a = 2.5$ and $b = 5.5$ and the latent variable for the FPC scores is $\xi \sim t(5)$: RB (green), PB with studentization (red), and naive PB with studentization (blue). Crosses $\times$ indicate coverage rates with $h_n$ selected by the proposed rule.

moments for $\xi$, the PB method with bootstrap studentization (3.16) performs well in most cases based on our rule of thumb.

### 3.5.2 Performance of bootstrap tests of projections

We now turn our attention to the testing problem discussed in Section 3.4. We investigate the empirical rejection rates of the bootstrap testing procedure when enforcing a null hypothesis of projection orthogonality, with bootstrap statistics from (3.21), or otherwise, with bootstrap statistics from (3.22).

The data generation for purposes of study are generally the same as considered in Section 3.5.1 with $\xi \sim \mathsf{N}(0, 1)$ and $a = 2.5$, with the exception that we modify the definition of the slope function $\beta$ to describe different hypotheses. For testing, the target predictors are considered as $\mathcal{X}_0 \equiv \{\phi_j\}_{j=1}^6$ based on the first six Fourier basis functions and we wish to assess the orthogonality of $\beta$ to the subspace spanned by $\mathcal{X}_0$ (i.e., $\Pi_{\mathcal{X}_0}\beta = 0$). Under the null hypothesis, the slope function is defined as $\beta^{H_0} \equiv \sum_{j\geq 6} W_{\beta,j}|\beta_j|\phi_j$, while the true data-generating slope is defined as $\beta^{H_1} \equiv (1-p)\beta^{H_0} + p\sum_{j=1}^6 W_{\beta,j}|\beta_j|\phi_j$ in terms of a proportion $p \in \{0, 0.1, \ldots, 0.9, 1\}$ for prescribing a sequence of alternative hypotheses; here $|\beta_j| = cj^{-b}$ holds with $c = 50$ and $b = 3.5$, and the value $p = 0$ renders the null hypothesis with the slope $\beta^{H_0}$.

We consider bootstrap tests of $H_0 : \Pi_{\mathcal{X}_0}\beta = 0$ based on a nominal size 5%. For each simulated dataset, 1000 bootstrap resamples are used to approximate the distribution of test statistics in (3.20). Truncation parameters $h_n$ and $g_n$ are again selected by the rule of thumb suggested in Section 3.5.1 based on $k_n$. Using 1000 simulation runs for each data generation scenario (level of $p$) and sample size $n$, we compute rejection rates by the proportion of times that an original test statistic exceeds the 95th percentile of bootstrap test distribution. The supplement [54] contains more details and findings over different sample sizes $n \in \{50, 200, 400, 600, 800, 1000\}$ as well as both test statistic forms from (3.20); we present results for $n = 50$ here with maximum or $L_\infty$ statistic form $W_{n,\max}$, as other results are qualitatively similar.

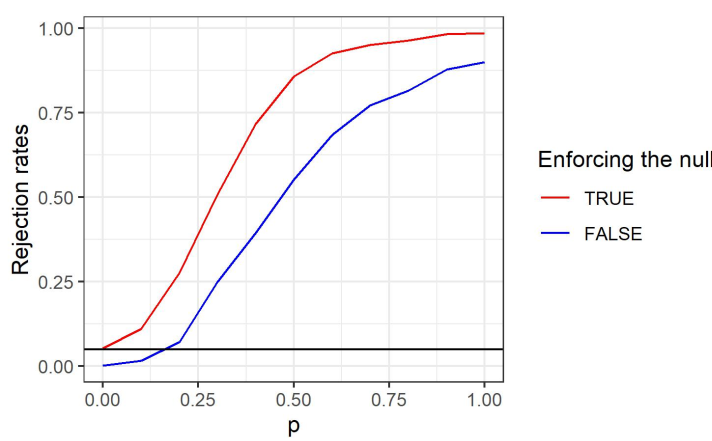


Figure 3.5: Empirical rejection rates (when $n = 50$) of the bootstrap testing procedure as the degree/proportion $p \in \{0, 0.25, 0.50, 0.75, 1\}$ of the alternative increases (only $p = 0$ corresponds to a true null hypothesis). The test may enforce the null hypothesis (red) or not (blue) in the bootstrap. The black horizontal line represents the nominal 0.05 size.

The resulting empirical rejection rates are summarized in Figure 3.5. As perhaps expected, the power of the test increases with the degree $p$ of how much the null hypothesis is violated, whether enforcing the null hypothesis (3.22) or not (3.21) in bootstrap. However, enforcing the null hypothesis maintains size better (i.e., when $p = 0$), which then also leads to slightly better power here. Another advantage to bootstrap enforcement of the null hypothesis is less sensitivity to choices of truncation parameters $h_n, g_n$. Results in the supplement [54] indicate that honoring the null hypothesis in bootstrap typically ensures good performance in testing as truncations $h_n, g_n$ are varied, which is not equally true for the bootstrap version that does not enforce the null (e.g., small $g_n$).

## 3.6 Real data analysis

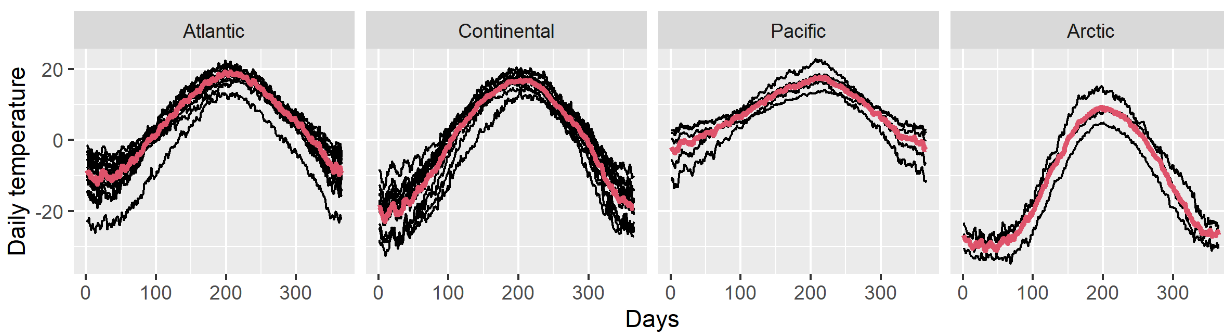


Figure 3.6: Daily temperature curves of locations in four different regions. Each black curve corresponds to the averaged in one location over 1964 to 1990, and the regional average curves are denoted in bold pink lines.

Bootstrap intervals and tests for projections are demonstrated with applications to Canadian weather data; another data application to medfly data is given in Section S4.2 of the supplement [54]. We analyze the Canadian weather dataset from the R package `fda` consisting of daily temperature and precipitation at 35 different locations in Canada [cf. 45]. The regressor $X_i$ is the daily temperature on each day averaged over 1960 to 1994, and the response $Y_i$ is the log of total annual precipitation with base 10. Here, $i$ indexes the $n = 35$ weather stations that record the temperatures and precipitations. The regressor curves $X_i$ are displayed in Figure 3.6, where the thicker lines represent the average for four different regions in Canada, namely, Atlantic, Continental, Pacific, and Arctic regions. The new predictors $\mathcal{X}_0 \equiv \{X_{0,l}\}_{l=1}^4$ under consideration for bootstrap inference are selected as these four average curves in each region as illustration. The centered observations are obtained as $X_i^c = X_i - \bar{X}$, $Y_i^c = Y_i - \bar{Y}$, and $X_{0,l}^c = X_{0,l} - \bar{X}$ before bootstrap inference. We will conduct inference of projections based on the proposed PB when the new (centered) daily temperatures are taken from $\mathcal{X}_0^c \equiv \{X_{0,l}^c\}_{l=1}^4$.

Each weather station is located in one of the four regions, where each region exhibits a different pattern as shown in Figure 3.6. This leads us to suspect the existence of different conditional variance of errors in FLRM (3.2). To investigate the heteroscedasticity, we estimate the variance from residuals for each region as $\hat{\sigma}_{r,k_n} = \left\{n_r^{-1} \sum_{i \in \mathcal{I}_r} (Y_i^c - \langle \hat{\beta}_{k_n}, X_i^c \rangle)^2 \right\}^{1/2}$, where $\mathcal{I}_r$ and $n_r$, respectively, denote the index set of and the number of location in the $r$th region. Here, the estimator $\hat{\beta}_{k_n}$ used for computing residuals is constructed from the combined data $\mathcal{D}_n^c \equiv \{(X_i^c, Y_i^c)\}_{i=1}^n$ over all four regions. As shown in Figure 3.1, homoscedastic error models seems implausible for this dataset. A similar conclusion is deduced from the residual plots given in Section 4.1 of the supplement.

Applying different bootstrap methods, the endpoints of 95% (symmetrized) confidence intervals for the (centered) projections $\{\langle \beta, X_{0,l}^c \rangle\}_{l=1}^4$ are given in Table 3.1. Here, the less consequential tuning parameter $k_n = 2$ was selected via repeated cross-validation, which minimizes prediction errors over estimates from $\hat{\beta}_{k_n}$, while $h_n = 2$ and $g_n = 2$ was then chosen by the rule of thumb suggested in Section 3.5.1. The supplement provides further results with different tuning parameter choices. As expected under possible heteroscedasticity and shown in Table 3.1, the results for residual bootstrap (RB) are quite different from those for PB, whether the latter is based on bootstrap studentization as in (3.16) (denoted as PB_std) or not as in (3.16) (denoted as PB). This distinction is also seen from a comparison of interval lengths in Table 3.1. Compared to the overall average, the Pacific region has the highest range of annual precipitation while the Continental region exhibits less precipitation with relatively narrow widths for both

Table 3.1: 95% symmetrized confidence intervals for projections $\{\langle\beta, X_{0,l}^c\rangle\}_{l=1}^4$ from RB, PB, and PB with studentization for Canadian weather dataset. The ratios of widths of RB intervals to widths of either PB or PB_std intervals are given in the parentheses.

| | RB | PB | PB_std |
|---|---|---|---|
| Atlantic | [ 0.06, 0.11] | [ 0.05, 0.12] (1.32) | [ 0.05, 0.12] (1.29) |
| Continental | [-0.19, -0.08] | [-0.18, -0.09] (0.84) | [-0.19, -0.08] (1.03) |
| Pacific | [ 0.18, 0.36] | [ 0.19, 0.35] (0.91) | [ 0.18, 0.36] (1.03) |
| Artic | [-0.49, -0.19] | [-0.58, -0.09] (1.61) | [-0.57, -0.10] (1.55) |

regions. The annual precipitation of the Atlantic and Arctic regions are respectively in either higher or lower range than the overall average, but with wider widths.

We apply our bootstrap testing procedure to this dataset for testing the null $H_0 : \Pi_{\mathcal{X}_0^c}\beta = 0$. Note that, because the regressors $\mathcal{X}_0^c \equiv \{X_{0,l}^c\}_{l=1}^4$ are centered by an overall average, this assessment is equivalent, in ANOVA fashion, to testing the null hypothesis of the equality of means across the four regions. The corresponding p-values are given in Table 3.2. All PB-based test statistics used strongly support that the slope function $\beta$ is not orthogonal to the space spanned by the predictors $\mathcal{X}_0^c$. That is, the data suggest that the true (uncentered) rainfall mean responses $\{\langle\beta, X_{0,l}\rangle\}_{l=1}^4$ are not equal at each regional mean curve and cannot be simultaneously equal to a common mean response $\langle\beta, \bar{X}\rangle$ at the global mean curve. This finding supports the region-wise PB intervals in Table 3.1.

Table 3.2: P-values for bootstrap testing of the null hypothesis $H_0 : \Pi_{\mathcal{X}_0^c}\beta = 0$ with different statistics for Canadian weather dataset.

| Enforcing the null | FALSE | TRUE |
|---|---|---|
| L2 | 0.000 | 0.001 |
| Max | 0.000 | 0.002 |

### 3.7 On establishing the CLT

We briefly outline of the proof of the CLT in Theorem 7; more technical details are provided in the supplement [54].

*Proof of Theorem 7.* The proof uses the following bias-variance decomposition of the functional principal component estimator $\hat{\beta}_{h_n}$ with respect to the true slope parameter $\beta$:

$$\hat{\beta}_{h_n} - \beta = b_n + \Gamma_{h_n}^{-1} U_n, \tag{3.24}$$

where, upon scaling $\sqrt{n/s_{h_n}(X_0)}$, the quantity $\Gamma_{h_n}^{-1} U_n$ determines the normal limit while a remainder/bias term $b_n \equiv \hat{\beta}_{h_n} - \beta - \Gamma_{h_n}^{-1} U_n$ converges to zero in probability. Above $U_n \equiv n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})(\varepsilon_i - \bar{\varepsilon})$ represents the cross-covariance between the regressors $\mathcal{X}_n \equiv \{X_i\}_{i=1}^{n}$ and the errors $\{\varepsilon_i\}_{i=1}^{n}$, with $\bar{X} \equiv n^{-1} \sum_{i=1}^{n} X_i$ and $\bar{\varepsilon} \equiv n^{-1} \sum_{i=1}^{n} \varepsilon_i$, and further $\Gamma_{h_n}^{-1} \equiv \sum_{j=1}^{h_n} \gamma_j^{-1} \pi_j$ denotes a truncated version of the inverse covariance operator $\Gamma^{-1} \equiv \sum_{j=1}^{\infty} \gamma_j^{-1} \pi_j$ with $\pi_j \equiv \phi_j \otimes \phi_j$ for integer $j \geq 1$. The supplement [54] shows that, as $n \to \infty$,

$$\mathsf{P}\left( \sqrt{\frac{n}{s_{h_n}(X_0)}} |\langle b_n, X_0 \rangle| > \eta \middle| X_0 \right) \xrightarrow{\mathsf{P}} 0 \tag{3.25}$$

holds for each $\eta > 0$. The distributional convergence of the term $\Gamma_{h_n}^{-1} U_n$ is stated in the following proposition, where the proof is deferred to the supplement [54].

**Proposition 15.** *Suppose that Conditions (A5)-(A7) hold. As $n \to \infty$, if $n^{-1} h_n^2 \to 0$ holds, then*

$$\sup_{y \in \mathbb{R}} \left| \mathsf{P}\left( \sqrt{\frac{n}{s_{h_n}(X_0)}} \langle \Gamma_{h_n}^{-1} U_n, X_0 \rangle \leq y \middle| X_0 \right) - \Phi(y) \right| \xrightarrow{\mathsf{P}} 0.$$

Theorem 7 then follows from (3.25) and Proposition 15 under the decomposition (3.38); see also Propositions S1-S3 in the supplement [54]. □

### 3.8 On proofs for the paired bootstrap

We sketch the proofs of Theorem 8 (consistency of the paired bootstrap) and Proposition 13 (inconsistency of naive paired bootstrap); further details appear in the supplement [54].

*Proof of Theorem 8.* To show bootstrap consistency, we consider a bootstrap-level decomposition, similar to (3.38), as

$$\hat{\beta}^*_{h_n} - \hat{\beta}_{g_n} = b^*_n + \Gamma^{-1}_{h_n}(U^*_n - \hat{U}_{n,g_n}) \tag{3.26}$$

where $b^*_n$ is a bias term, $\hat{U}_{n,g_n}$ is the bias correction from (3.7), and $U^*_n \equiv n^{-1}\sum_{i=1}^n (X^*_i - \bar{X}^*)(\varepsilon^*_{i,g_n} - \bar{\varepsilon}^*_{g_n})$ denotes the sample cross covariance between the bootstrap regressors $\{X^*_i\}_{i=1}^n$ and the bootrstrap errors $\{\varepsilon^*_{i,g_n}\}_{i=1}^n$, where $\bar{X}^* \equiv n^{-1}\sum_{i=1}^n X^*_i$ and $\bar{\varepsilon}^*_{g_n} \equiv n^{-1}\sum_{i=1}^n \varepsilon^*_{i,g_n}$ from $\varepsilon^*_{i,g_n} \equiv Y^*_i - \langle \hat{\beta}_{g_n}, X^*_i\rangle$. Proposition 16 shows that, upon scaling, the distribution of $\Gamma^{-1}_{h_n}(U^*_n - \hat{U}_{n,g_n})$ under bootstrap probability $\mathsf{P}^*(\cdot|X_0)$ converges to a standard normal distribution.

**Proposition 16.** *Suppose that Conditions (A1)-(A7) hold, and that* $\|\hat{\beta}_{g_n} - \beta\| \xrightarrow{\mathsf{P}} 0$ *and* $n^{-1}\sum_{j=1}^{h_n} \lambda_j^{-1} = O(1)$ *as* $n \to \infty$*. Then, as* $n \to \infty$,

$$\sup_{y\in\mathbb{R}} \left| \mathsf{P}^*\left( \sqrt{\frac{n}{s_{h_n}(X_0)}} \langle \Gamma^{-1}_{h_n}(U^*_n - \hat{U}_{n,g_n}), X_0\rangle \le y \middle| X_0 \right) - \Phi(y) \right| \xrightarrow{\mathsf{P}} 0.$$

The supplement [54] then establishes that a scaled projection involving $b^*_n \equiv \hat{\beta}^*_{h_n} - \hat{\beta}_{g_n} - \Gamma^{-1}_{h_n}(U^*_n - \hat{U}_{n,g_n})$ from (3.26) converges to zero in bootstrap probability $\mathsf{P}^*(\cdot|X_0)$, namely,

$$\mathsf{P}^*\left( \sqrt{\frac{n}{s_{h_n}(X_0)}} |\langle b^*_n, X_0\rangle| > \eta \middle| X_0 \right) \xrightarrow{\mathsf{P}} 0, \tag{3.27}$$

as $n \to \infty$, for each $\eta > 0$. Using a subsequence argument (cf. [4], Theorem 20.5) for bootstrap distributions along with Slutsky's theorem, Theorem 8 then follows from (3.27) in combination with Proposition 16 and (3.26); see also Propositions S5-S9 in the supplement [54].

□

*Proof of Proposition 16.* We write $Z^*_{i,n} = \langle X^*_i \varepsilon^*_{i,g_n} - \tilde{U}_{n,g_n}, \Gamma^{-1}_{h_n} X_0\rangle$ with $\tilde{U}_{n,g_n} \equiv n^{-1}\sum_{i=1}^n X^*_i \varepsilon^*_{i,g_n}$ so that $\mathsf{E}^*[Z^*_{i,n}|X_0] = 0$ and

$$\begin{aligned} &\sqrt{\frac{n}{s_{h_n}(X_0)}} \langle \Gamma^{-1}_{h_n}(U^*_n - \hat{U}_{n,g_n}), X_0\rangle \\ =&\{n s_{h_n}(X_0)\}^{-1/2} \sum_{i=1}^n Z^*_{i,n} - \sqrt{\frac{n}{s_{h_n}(X_0)}} \langle \bar{X}^* \bar{\varepsilon}^*_{g_n} - \bar{X}\bar{\hat{\varepsilon}}_{g_n}, \Gamma^{-1}_{h_n} X_0\rangle, \end{aligned} \tag{3.28}$$

where $\bar{\hat{\varepsilon}}_{g_n} \equiv n^{-1}\sum_{i=1}^n \hat{\varepsilon}_{i,g_n}$ with $\hat{\varepsilon}_{i,g_n} \equiv Y_i - \langle \hat{\beta}_{g_n}, X_i\rangle$. Conditional on $X_0$, define the bootstrap variance $\hat{v}_n^2 \equiv \sum_{i=1}^n \mathsf{E}^*[Z_{i,n}^{*2}|X_0]$ and a bootstrap version of the Lindeberg condition as $\hat{\mathcal{L}}_n \equiv \hat{v}_n^{-2}\sum_{i=1}^n \mathsf{E}^*[Z_{i,n}^{*2}\mathbb{I}(|Z_{i,n}^*| > \tau\hat{v}_n)|X_0]$ for given $\tau > 0$. To establish a bootstrap CLT, we use assertions (3.29)-(3.32) below, proved in Section S2.4 of the supplement [54]: as $n \to \infty$,

$$\mathsf{E}^*\left[\left\{\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle \bar{X}^*\bar{\varepsilon}^* - \bar{X}\bar{\hat{\varepsilon}}, \Gamma_{h_n}^{-1}X_0\rangle\right\}^2 \Big| X_0\right] \xrightarrow{\mathsf{P}} 0; \tag{3.29}$$

$$\frac{n^{-1}\hat{v}_n^2}{s_{h_n}(X_0)} \xrightarrow{\mathsf{P}} 1; \tag{3.30}$$

$$\mathsf{E}^*\left[\left(\hat{v}_n^{-1}\max_{1\le i\le n}|Z_{i,n}^*|\right)^4 \Big| X_0\right] \xrightarrow{\mathsf{P}} 0; \tag{3.31}$$

$$\mathsf{E}^*\left[\left|\frac{n^{-1}\sum_{i=1}^n Z_{i,n}^{*2}}{s_{h_n}(X_0)} - 1\right| \Big| X_0\right] \xrightarrow{\mathsf{P}} 0. \tag{3.32}$$

Results in (3.30) and (3.32) further yield that

$$\mathsf{E}^*\left[\left|\hat{v}_n^{-2}\sum_{i=1}^n Z_{i,n}^{*2} - 1\right| \Big| X_0\right] \xrightarrow{\mathsf{P}} 0. \tag{3.33}$$

We next write $A_n^* \equiv \hat{v}_n^{-2}\sum_{i=1}^n Z_{i,n}^{*2}$, $B_n^* \equiv \hat{v}_n^{-2}\sum_{i=1}^n Z_{i,n}^{*2}\mathbb{I}(|Z_{i,n}^*| > \eta\hat{v}_n)$, and $C_n^* \equiv \hat{v}_n^{-1}\max_{1\le i\le n}|Z_{i,n}^*|$. Using a subsequence argument on an almost sure set and conditional on $X_0$ (cf. [4], Theorem 20.5), one can verify that (i) $B_n^* \to 0$ holds in bootstrap probability $\mathsf{P}^*(\cdot|X_0)$, because $B_n^* \le A_n^*\mathbb{I}(C_n^* > \tau) \to 1\cdot 0$ in bootstrap probability from (3.31) and (3.33); and also that (ii) $\{B_n^*\}$ is uniformly integrable with respect to $\mathsf{P}^*(\cdot|X_0)$, because $B_n^* \le A_n^*$ holds and $\{A_n^*\}$ is likewise uniformly integrable by (3.33). When (i)-(ii) hold, then $\hat{\mathcal{L}}_n = \mathsf{E}^*(B_n^*|X_0) \to \mathsf{E}^*(0|X_0) = 0$ follows along the same subsequence almost surely. Hence, we conclude that the Lindeberg term $\hat{\mathcal{L}}_n$ converges to zero in probability (cf. [2], Theorem 9.5.1). This fact, together with (3.29) and the expansion in (3.28), yields

$$\sup_{y\in\mathbb{R}}\left|\mathsf{P}^*\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle \Gamma_{h_n}^{-1}(U_n^* - \hat{U}_{n,g_n}), X_0\rangle \le y \Big| X_0\right) - \Phi(y)\right| \xrightarrow{\mathsf{P}} 0$$

by Polya's theorem and the continuity of the standard normal distribution function $\Phi$. □

*Proof of Proposition 13.* By Propositions S11-S15 and Lemma S39 in the supplement [54], the naive bootstrap construction $T_{n,naive}^*(X_0)$ can be written as

$T^*_{n,naive}(X_0) = T^*_{n,\hat{s}}(X_0) + A^*_n + B_n + C_n$, where $T^*_{n,\hat{s}}(X_0)$ is the PB quantity from (3.14); $A^*_n \equiv A^*_n(X_0)$ is a bootstrap error term that converges to zero in bootstrap probability if $n^{-1/2}h_n^4(\log h_n)^{7/2} \to 0$; $B_n \equiv B_n(X_0)$ represents a bias-type term that does not depend on the bootstrap sample and satisfies $\sup_{y\in\mathbb{R}} |\mathsf{P}(B_n \leq y|X_0) - \Phi(y/\sigma(\tau))| \xrightarrow{\mathsf{P}} 0$ with limit variance $\sigma^2(\tau)$ from (3.18); and $C_n \equiv C_n(X_0)$ is a negligible term that converges to zero if $n^{-1/2}h_n^{9/2}(\log h_n)^6 \to 0$. By writing $D_n \equiv B_n + C_n$ and applying the triangle inequality, we find

$$\left| \sup_{y\in\mathbb{R}} |\mathsf{P}^*(T^*_{n,naive}(X_0) \leq y|X_0) - \mathsf{P}(T_n(X_0) \leq y|X_0)| - \sup_{y\in\mathbb{R}} |\Phi(y - D_n) - \Phi(y)| \right| \xrightarrow{\mathsf{P}} 0,$$

using that $\sup_{y\in\mathbb{R}} |\mathsf{P}^*(T^*_{n,\hat{s}}(X_0) + A^*_n \leq y|X_0) - \Phi(y)| \xrightarrow{\mathsf{P}} 0$ by Theorem 8 with Proposition S11 in [54] and that $\sup_{y\in\mathbb{R}} |\mathsf{P}(T_n(X_0) \leq y|X_0) - \Phi(y)| \xrightarrow{\mathsf{P}} 0$ by Theorem 7. By the continuous mapping theorem/embedding theorem, we then have

$$\Phi(y - D_n) - \Phi(y) \xrightarrow{\mathsf{d}} \Phi\Big(y + \sigma(\tau)Z\Big) - \Phi(y), \quad y \in \mathbb{R},$$

as elements in $\mathbb{D}$, based on $D_n \equiv B_n + C_n \xrightarrow{\mathsf{d}} -\sigma(\tau)Z$ for a standard normal variable $Z$. The convergence in Proposition 13 then follows (cf. [4]). □

## 3.9 Technical details: central limit theorem

This section contains the technical results to prove the central limit theorem (CLT) for projection estimator under heteroscedasticity provided in Section 3.3.1 of the main text. First, some preliminary lemmas from functional calculus are described in Section 3.9.1. In Section 3.9.2, we next prove the consistency of the estimator $\hat{\beta}_{h_n}$ for the slope function $\beta$. Section 3.9.3 then completes the proofs of the CLT described in Theorem 7 of the main text.

### 3.9.1 Prelimimaries: functional calculus

We introduce some preliminary results from the perturbation theory or functional calculus in functional analysis. Such techniques are now common in functional data analysis literature. We refer to [21], Chapter VII, [23], Chapter I, or [33], Chapter 5. Since we reflect centering by averages in the estimation unlike the existing works such as [12, 53], the technical lemmas and their proofs are slightly modified.

Write $\|\cdot\|_\infty$ and $\|\cdot\|_{HS}$ for the operator supremum norm and Hilbert-Schmidt norm respectively. Let $\mathcal{B}_j = \{z \in \mathbb{C} : |z-\gamma_j| \le \delta_j/2\}$ be the oriented circle in the complex plane $\mathbb{C}$ and set $\mathcal{C}_{h_n} = \sum_{j=1}^{h_n} \mathcal{B}_j$ to define the contour integral for operator-valued functions. By the theory from functional calculus (for the bounded linear operators) or perturbation theory, we see that

$$\Pi_{h_n} = \sum_{j=1}^{h_n} \pi_j = \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}(zI-\Gamma)^{-1}dz = \frac{1}{2\pi\iota}\int_{\mathcal{C}_{h_n}}(zI-\Gamma)^{-1}dz, \tag{3.34}$$

$$\Gamma_{h_n}^{-1} = \sum_{j=1}^{h_n} \gamma_j^{-1}\pi_j = \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}z^{-1}(zI-\Gamma)^{-1}dz = \frac{1}{2\pi\iota}\int_{\mathcal{C}_{h_n}}z^{-1}(zI-\Gamma)^{-1}dz, \tag{3.35}$$

where $\pi_j = \phi_j \otimes \phi_j = \frac{1}{2\pi\iota}\int (zI-\Gamma)^{-1}dz$ denotes the Riesz projection of $\Gamma$ corresponding to the $j$-th eigenvalue $\gamma_j$, which is the projection operator onto the $j$-th eigenfunction $\phi_j$. One can also get the empirical counterparts $\hat{\Pi}_{h_n}$ and $\hat{\Gamma}_{h_n}^{-1}$ to the above contour integral forms from the sample covariance operator $\hat{\Gamma}_n$ with the corresponding random contours $\hat{\mathcal{B}}_j = \{z \in \mathbb{C} : |z-\hat{\gamma}_j| \le \hat{\delta}_j/2\}$ and $\hat{\mathcal{C}}_{h_n} = \bigcup_{j=1}^{h_n}\hat{\mathcal{B}}_j$. For further purposes, we use the following notations:

$$G_n(z) = (zI-\Gamma)^{-1/2}(\hat{\Gamma}_n-\Gamma)(zI-\Gamma)^{-1/2};$$

$$K_n(z) = (zI-\Gamma)^{1/2}(zI-\hat{\Gamma}_n)^{-1}(zI-\Gamma)^{1/2};$$

$$\mathcal{E}_j = (\|G_n(z)\|_\infty < 1/2, \forall z \in \mathcal{B}_j);$$

$$\mathcal{A}_{h_n} = \{\forall j \in \{1,\dots,h_n\}, |\hat{\gamma}_j-\gamma_j| < \delta_j/2\}.$$

The following lemmas originally come from [12] and can be generalized to the case when covariance estimators $\hat{\Gamma}_n$ and $\hat{\Delta}_n$ are centered and when the error variances are heterogeneous. Lemmas 11, 12, 14, 16, 17 can be proved in the same way as in [12, 53]. Lemma 13 is a preliminary result to prove Lemma 15 and requires a slight modification due to centering in estimation. In Lemma 15, we added new results, which are proved from the same argument as the proof of Lemma 3 in [12].

**Lemma 11.** *Suppose that $\gamma_j$ is a convex function of $j$ (which implies that $\delta_j = \gamma_j - \gamma_{j+1}$ is decreasing) at least for sufficiently large $j$. Suppose the Condition (A3) holds. Then, for* *sufficiently large $j, k \in \mathbb{N}$ with $k > j$, we have*

$$j\gamma_j \geq k\gamma_k, \quad \frac{\gamma_j}{\gamma_j - \gamma_k} \leq \frac{k}{k-j}, \quad \textit{and} \quad \sum_{j \geq k} \gamma_j \leq (k+1)\gamma_k.$$

**Lemma 12.** *Under the same assumptions of Lemma 11, we have that*

$$\sum_{l \neq j} \frac{\gamma_l}{|\gamma_l - \gamma_j|} \leq Cj \log j$$

*for sufficiently large $j \in \mathbb{N}$.*

**Lemma 13.** *Suppose that Condition (A2) holds. We then have that*

$$\sup_{j,k \in \mathbb{N}} \frac{\mathsf{E}[\langle (\hat{\Gamma}_n - \Gamma)\phi_j, \phi_k \rangle^2]}{\gamma_j \gamma_k} \leq \frac{C_1}{n} + \frac{C_2}{n^2}.$$

*Proof.* Note that

$$\begin{aligned}
\langle (\hat{\Gamma}_n - \Gamma)\phi_j, \phi_k \rangle^2 &= \{\langle (\tilde{\Gamma}_n - \Gamma)\phi_j, \phi_k \rangle - \langle \bar{X}^{\otimes 2}\phi_j, \phi_k \rangle\}^2 \\
&\leq 2\langle (\tilde{\Gamma}_n - \Gamma)\phi_j, \phi_k \rangle^2 + 2\langle \bar{X}, \phi_j \rangle^2 \langle \bar{X}, \phi_k \rangle^2,
\end{aligned}$$

where $\tilde{\Gamma}_n \equiv n^{-1} \sum_{i=1}^n X_i^{\otimes 2}$.

To check the upper bound for the first term in the above display, we observe the following identity:

$$\begin{aligned}
&\langle (\tilde{\Gamma}_n - \Gamma)\phi_j, \phi_k \rangle^2 \\
=&\left\{ n^{-1} \sum_{i=1}^n \langle X_i, \phi_j \rangle \langle X_i, \phi_k \rangle - \langle \Gamma\phi_j, \phi_k \rangle \right\}^2 \\
=&n^{-2} \left( \sum_{i=1}^n \langle X_i, \phi_j \rangle \langle X_i, \phi_k \rangle \right)^2 - \frac{2}{n} \sum_{i=1}^n \langle X_i, \phi_j \rangle \langle X_i, \phi_k \rangle \langle \Gamma\phi_j, \phi_k \rangle + \langle \Gamma\phi_j, \phi_k \rangle^2 \\
=&n^{-2} \sum_{i=1}^n \langle X_i, \phi_j \rangle^2 \langle X_i, \phi_k \rangle^2 + n^{-2} \sum_{i \neq i'} \langle X_i, \phi_j \rangle \langle X_i, \phi_k \rangle \langle X_{i'}, \phi_j \rangle \langle X_{i'}, \phi_k \rangle \\
&- \frac{2}{n} \sum_{i=1}^n \langle X_i, \phi_j \rangle \langle X_i, \phi_k \rangle \langle \Gamma\phi_j, \phi_k \rangle + \langle \Gamma\phi_j, \phi_k \rangle^2.
\end{aligned}$$

Since $X_i$ and $X_{i'}$ are independent and

$$\begin{aligned}
\mathsf{E}[\langle X_i, \phi_j \rangle \langle X_i, \phi_k \rangle] = \mathsf{E}[\langle X^{\otimes 2}\phi_j, \phi_k \rangle] &= \langle \mathsf{E}[X^{\otimes 2}]\phi_j, \phi_k \rangle \\
&= \langle \Gamma\phi_j, \phi_k \rangle
\end{aligned}$$

we have that

$$\begin{aligned}
&\mathsf{E}[\langle(\tilde{\Gamma}_n-\Gamma)\phi_j,\phi_k\rangle^2] \\
=&n^{-1}\mathsf{E}[\langle X,\phi_j\rangle^2\langle X,\phi_k\rangle^2]+\frac{n^2-n}{n^2}\langle\Gamma\phi_j,\phi_k\rangle^2-2\langle\Gamma\phi_j,\phi_k\rangle^2+\langle\Gamma\phi_j,\phi_k\rangle^2 \\
=&n^{-1}\left(\mathsf{E}[\langle X,\phi_j\rangle^2\langle X,\phi_k\rangle^2]-\langle\Gamma\phi_j,\phi_k\rangle^2\right) \\
\leq&n^{-1}\gamma_j\gamma_k\mathsf{E}[\xi_j^2\xi_k^2]\leq\frac{\gamma_j\gamma_k}{n}\sqrt{\mathsf{E}[\xi_j^4]}\sqrt{\mathsf{E}[\xi_k^4]}.
\end{aligned}$$

Since $\sup_{j\in\mathbb{N}}\mathsf{E}[\xi_j^4]<\infty$ by assumption, we see that

$$\sup_{j,k\in\mathbb{N}}\frac{n}{\gamma_j\gamma_k}\mathsf{E}[\langle(\tilde{\Gamma}_n-\Gamma)\phi_j,\phi_k\rangle^2]<\infty.$$

To investigate the next term, we first consider the case of $j\neq k$. Then, we have

$$\begin{aligned}
&\mathsf{E}[\langle\bar{X},\phi_j\rangle^2\langle\bar{X},\phi_k\rangle^2] \\
=&n^{-4}\sum_{i\neq i'}\mathsf{E}[\langle X_i,\phi_j\rangle^2\langle X_{i'},\phi_k\rangle^2]+n^{-4}\sum_{i=1}^n\mathsf{E}[\langle X_i,\phi_j\rangle^2\langle X_i,\phi_k\rangle^2] \\
=&n^{-4}\{(n^2-n)\gamma_j\gamma_k+n\mathsf{E}[\langle X,\phi_j\rangle^2\langle X,\phi_k\rangle^2]\} \\
\leq&n^{-4}\left\{(n^2-n)\gamma_j\gamma_k+n\gamma_j\gamma_k\sqrt{\mathsf{E}[\xi_j^4]}\sqrt{\mathsf{E}[\xi_k^4]}\right\} \\
\leq&n^{-4}\{(n^2-n)\gamma_j\gamma_k+Cn\gamma_j\gamma_k\} \\
\leq&C\frac{\gamma_j\gamma_k}{n^2}
\end{aligned}$$

since $\sup_{j\in\mathbb{N}}\mathsf{E}[\xi_j^4]<\infty$ by assumption. Similarly, if $j=k$, then

$$\begin{aligned}
&\mathsf{E}[\langle\bar{X},\phi_j\rangle^4] \\
=&3n^{-4}\sum_{i\neq i'}\mathsf{E}[\langle X_i,\phi_j\rangle^2\langle X_{i'},\phi_j\rangle^2]+n^{-4}\sum_{i=1}^n\mathsf{E}[\langle X_i,\phi_j\rangle^4] \\
\leq&3n^{-4}\{(n^2-n)\gamma_j^2+n\gamma_j^2\} \\
\leq&C\frac{\gamma_j\gamma_k}{n^2}.
\end{aligned}$$

We thus have that

$$\sup_{j,k\in\mathbb{N}}\frac{n^2}{\gamma_j\gamma_k}\mathsf{E}[\langle\bar{X},\phi_j\rangle^2\langle\bar{X},\phi_k\rangle^2]<\infty.$$

□

**Lemma 14** (Second resolvent identity)**.** *The difference between the resolvents of* $\hat{\Gamma}_n$ *and* $\Gamma$ *can be written as*

$$(zI - \hat{\Gamma}_n)^{-1} - (zI - \Gamma)^{-1} = (zI - \hat{\Gamma}_n)^{-1}(\hat{\Gamma}_n - \Gamma)(zI - \Gamma)^{-1}$$
$$= (zI - \Gamma)^{-1}(\hat{\Gamma}_n - \Gamma)(zI - \hat{\Gamma}_n)^{-1},$$

*and hence,*

$$(zI - \hat{\Gamma}_n)^{-1} - (zI - \Gamma)^{-1} = (zI - \Gamma)^{-1/2} K_n(z) G_n(z)(zI - \Gamma)^{-1/2}.$$

**Lemma 15.** *Suppose that* $\gamma_j$ *is a convex function of* $j$ *at least for sufficiently large* $j$*. Also, suppose that* $\sup_{j\in\mathbb{N}} \mathsf{E}[\xi_j^4] < \infty$*. Suppose that Conditions (A2)-(A3) hold. Then, for sufficiently large* $j$*, we have the following.*

1. $\mathsf{E}\left[\sup_{z\in\mathcal{B}_j} \|G_n(z)\|_\infty^2\right] \le Cn^{-1}(j\log j)^2$;
2. $\mathsf{E}\left[\sup_{z\in\mathcal{B}_j} \left\|(zI - \Gamma)^{-1/2}X\right\|^2\right] \le Cj\log j$;
3. $\mathsf{E}\left[\sup_{z\in\mathcal{B}_j} \|(zI - \Gamma)^{-1/2}X\|^4\right] \le C(j\log j)^2$; *and*
4. $\mathsf{E}\left[\sup_{z\in\mathcal{B}_j} \left\|(zI - \Gamma)^{-1/2}U_n\right\|^2\right] \le Cn^{-1}\delta_j^{-1}$.

*Proof.* The first and second assertions are proved in [CMS]. For the third part, with a similar argument to the proof of Lemma 3 in [CMS], we have

$$\sup_{z\in\mathcal{B}_j} \|(zI - \Gamma)^{-1/2}X\|^4 \le 4\left(\sum_{k\ne j}\frac{\gamma_k\xi_k^2}{|\gamma_j - \gamma_k|} + \frac{\gamma_j\xi_j^2}{\delta_j}\right)^2$$
$$= 4\left\{\sum_{l\ne k, j\ne k}\frac{\gamma_l\gamma_k\xi_l^2\xi_k^2}{|\gamma_j - \gamma_l||\gamma_j - \gamma_k|} + \sum_{k\ne j}\frac{\gamma_j\gamma_k\xi_j^2\xi_k^2}{\delta_j|\gamma_j - \gamma_k|} + \frac{\gamma_j^2\xi_j^4}{\delta_j^2}\right\},$$

which implies that

$$\mathsf{E}\left[\sup_{z\in\mathcal{B}_j} \|(zI - \Gamma)^{-1/2}X\|^4\right] \le C\left\{\sum_{l\ne k, j\ne k}\frac{\gamma_l\gamma_k}{|\gamma_j - \gamma_l||\gamma_j - \gamma_k|} + \sum_{k\ne j}\frac{\gamma_j\gamma_k}{\delta_j|\gamma_j - \gamma_k|} + \frac{\gamma_j^2}{\delta_j^2}\right\}$$
$$\le C(j\log j)^2.$$

For the last part, we note that

$$\begin{aligned}\|(zI-\Gamma)^{-1/2}U_n\|^2 &= \|(zI-\Gamma)^{-1/2}\tilde{U}_n - \bar{\varepsilon}(zI-\Gamma)^{-1/2}\bar{X}\|^2 \\ &\leq 2\|(zI-\Gamma)^{-1/2}\tilde{U}_n\|^2 + 2\bar{\varepsilon}^2\|(zI-\Gamma)^{-1/2}\bar{X}\|^2, \end{aligned} \tag{3.36}$$

where $\tilde{U}_n \equiv n^{-1}\sum_{i=1}^n \varepsilon_i X_i$.

To bound the first term in (3.36), note that

$$\begin{aligned}&\|(zI-\Gamma)^{-1/2}\tilde{U}_n\|^2 \\ =&n^{-2}\sum_{i=1}^n \|(zI-\Gamma)^{-1/2}X_i\|^2\varepsilon_i^2 \\ &+ n^{-2}\sum_{i\neq i'}\langle (zI-\Gamma)^{-1/2}X_i, (zI-\Gamma)^{-1/2}X_{i'}\rangle \varepsilon_i\varepsilon_{i'},\end{aligned}$$

and

$$\begin{aligned}&\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\tilde{U}_n\|^2 \\ \leq& n^{-2}\sum_{i=1}^n\left(\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2\right)\varepsilon_i^2 \\ &+ n^{-2}\sum_{i\neq i'}\left(\sup_{z\in\mathcal{B}_j}\langle (zI-\Gamma)^{-1/2}X_i, (zI-\Gamma)^{-1/2}X_{i'}\rangle\right)\varepsilon_i\varepsilon_{i'}.\end{aligned}$$

This implies that

$$\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\tilde{U}_n\|^2\right] \leq n^{-1}\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^2\varepsilon^2\right]$$

since

$$\begin{aligned}&\mathsf{E}\left[\left(\sup_{z\in\mathcal{B}_j}\langle (zI-\Gamma)^{-1/2}X_i, (zI-\Gamma)^{-1/2}X_{i'}\rangle\right)\varepsilon_i\varepsilon_{i'}\right] \\ =&\mathsf{E}\left[\left(\sup_{z\in\mathcal{B}_j}\langle (zI-\Gamma)^{-1/2}X_i, (zI-\Gamma)^{-1/2}X_{i'}\rangle\right)\mathsf{E}[\varepsilon_i\varepsilon_{i'}|\mathcal{X}_n]\right] \\ =&0\end{aligned}$$

from $\mathsf{E}[\varepsilon_i\varepsilon_{i'}|\mathcal{X}_n] = \mathsf{E}[\varepsilon_i|\mathcal{X}_n]\mathsf{E}[\varepsilon_{i'}|\mathcal{X}_n] = \mathsf{E}[\varepsilon_i|X_i]\mathsf{E}[\varepsilon_{i'}|X_{i'}] = 0$. By Equation (5.3) in [33], for $z \in \mathcal{B}_j$,

$$\|(zI-\Gamma)^{-1/2}\|_\infty = \left(\min_{l\in\mathbb{N}}|z-\gamma_l|^{1/2}\right)^{-1} = |z-\gamma_j|^{-1/2} = (\delta_j/2)^{-1/2}.$$

This implies that

$$\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^2\varepsilon^2\right]\le \mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\|_\infty^2\|X\varepsilon\|^2\right]$$
$$\le 2\delta_j^{-1}\mathsf{E}[\|X\varepsilon\|^2]=2\mathrm{tr}(\mathsf{var}[X\varepsilon])\delta_j^{-1}.$$

Meanwhile, to find an upper bound for the second term in (3.36), note that

$$\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\bar{X}\|^2\bar{\varepsilon}^2\right]\le \sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\|^2\mathsf{E}[\|\bar{X}\bar{\varepsilon}\|^2]=2\delta_j^{-1}\mathsf{E}[\|\bar{X}\bar{\varepsilon}\|^2].$$

We see that

$$\|\bar{X}\bar{\varepsilon}\|^2=n^{-4}\left\|\sum_{i,i_0}X_i\varepsilon_{i_0}\right\|^2=n^{-4}\sum_{i,i',i_0,i_0'}\langle X_i,X_i'\rangle\varepsilon_{i_0}\varepsilon_{i_0'}.$$

Since $\mathsf{E}[\varepsilon_i|X_i]=0$, we have that

$$\mathsf{E}[\|\bar{X}\bar{\varepsilon}\|^2]=n^{-4}\sum_{i,i',i_0}\mathsf{E}[\langle X_i,X_i'\rangle\varepsilon_{i_0}^2].$$

Note that $\mathsf{E}[\langle X_i,X_i'\rangle\varepsilon_{i_0}^2]\le\mathsf{E}[\|X_i\varepsilon_{i_0}\|\|X_i'\varepsilon_{i_0}\|]\le\mathsf{E}[\|X_i\varepsilon_{i_0}\|^2]^{1/2}\mathsf{E}[\|X_i'\varepsilon_{i_0}\|^2]^{1/2}$ by Cauchy-Schwarz inequality. If $i\ne i_0$, $\mathsf{E}[\|X_i\varepsilon_{i_0}\|^2]=\mathsf{E}[\|X_i\|^2]\mathsf{E}[\varepsilon_{i_0}^2]<\infty$. If $i=i_0$, $\mathsf{E}[\|X_i\varepsilon_i\|^2]=\mathsf{E}[\|X\varepsilon\|^2]=\mathrm{tr}(\mathsf{var}[X\varepsilon])<\infty$. This implies that

$$\mathsf{E}\left[\|\bar{X}\bar{\varepsilon}\|^2\right]\le Cn^{-1}, \tag{3.37}$$

and hence,

$$\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\bar{X}\|^2\bar{\varepsilon}^2\right]\le Cn^{-1}\delta_j^{-1}.$$

□

**Lemma 16.** *Suppose the same assumptions of Lemma 15. We have that* $\sup_{z\in\mathcal{B}_j}\|K_n(z)\|_\infty\mathbb{I}_{\mathcal{E}_j}\le 2$ *almost surely and* $\mathsf{P}(\mathcal{E}_j^c)\le Cn^{-1/2}j\log j$.

**Lemma 17.**

1. *We observe that*

$$\hat{\Pi}_{h_n} - \Pi_{h_n} = \frac{1}{2\pi\iota}\int_{\mathcal{C}_{h_n}} \{(zI - \hat{\Gamma}_n)^{-1} - (zI - \Gamma)^{-1}\}dz + r_{1n}\mathbb{I}_{\mathcal{A}^c_{h_n}},$$
$$\hat{\Gamma}^{-1}_{h_n} - \Gamma^{-1}_{h_n} = \frac{1}{2\pi\iota}\int_{\mathcal{C}_{h_n}} z^{-1}\{(zI - \hat{\Gamma}_n)^{-1} - (zI - \Gamma)^{-1}\}dz + r_{2n}\mathbb{I}_{\mathcal{A}^c_{h_n}},$$

*where*

$$r_{1n} = \hat{\Pi}_{h_n} - \frac{1}{2\pi\iota}\int_{\mathcal{C}_{h_n}} (zI - \hat{\Gamma}_n)^{-1}dz,$$
$$r_{2n} = \hat{\Gamma}^{-1}_{h_n} - \frac{1}{2\pi\iota}\int_{\mathcal{C}_{h_n}} z^{-1}(zI - \hat{\Gamma}_n)^{-1}dz.$$

2. *Suppose that $\gamma_j$ is a convex function of $j$ at least for sufficiently large $j$ and that $\sup_{j\in\mathbb{N}} \mathsf{E}[\xi_j^4] < \infty$. We then have that*

$$\mathsf{P}(\mathcal{A}^c_{h_n}) \leq C_1 n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2} + C_2 n^{-1/2}\sum_{j=1}^{h_n} j\log j.$$

*Remark* 14. Instead of the fourth result in Lemma 15, one can derive a more specific upper bound of $\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\left\|(zI-\Gamma)^{-1/2}U_n\right\|^2\right]$ under some error structures. For example, if either $\mathsf{E}[\varepsilon^4] < \infty$ or $\mathsf{E}[\varepsilon^2|X] = \sum_{j=1}^\infty \rho_j^2\langle X, \phi_j\rangle^2$ for some $\{\rho_j\}_{j=1}^\infty$ with $\sum_{j=1}^\infty \gamma_j\rho_j^2 < \infty$, the upper bound $Cj\log j$ can be obtained. This determines the convergence rate on $h_n$ for the bias term related to $\langle(\hat{\Gamma}^{-1}_{h_n} - \Gamma^{-1}_{h_n})U_n, X_0\rangle$ (cf. Proposition 18) and hence the rate for Theorem 7. This allows us to the same growth rate $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}(j\log j)^2 \to 0$ as the ones used for the CLT under homoscedasticity (cf. [12, 53]).

Some preliminary lemmas related to centering by sample means $\bar{X} \equiv n^{-1}\sum_{i=1}^n X_i$ and $\bar{\varepsilon} \equiv n^{-1}\sum_{i=1}^n \varepsilon_i$ are provided first.

**Lemma 18.** *We have that $\mathsf{E}[\|\bar{X}\bar{\varepsilon}\|^2] \leq Cn^{-2}$.*

*Proof.* Note that

$$\bar{X}\bar{\varepsilon} = n^{-2}\sum_{i=1}^n X_i\varepsilon_i + n^{-2}\sum_{i\neq i'} X_i\varepsilon_{i'}$$

so that

$$\|\bar{X}\bar{\varepsilon}\|^2 \le 2n^{-4}\left\|\sum_{i=1}^{n} X_i\varepsilon_i\right\|^2 + 2n^{-4}\left\|\sum_{i\neq i'} X_i\varepsilon_{i'}\right\|^2$$
$$= 2n^{-4}\left(\sum_{i=1}^{n}\|X_i\varepsilon_i\|^2 + \sum_{i\neq i'}\langle X_i\varepsilon_i, X_{i'}\varepsilon_{i'}\rangle\right) + 2n^{-4}\sum_{i\neq i', i_0\neq i_0'}\langle X_i\varepsilon_{i'}, X_{i_0}\varepsilon_{i_0'}\rangle.$$

We first see that $\mathsf{E}[\|X_i\varepsilon_i\|^2] = \mathsf{E}[\|X\varepsilon\|^2] = \mathrm{tr}(\Lambda) < \infty$ and $\mathsf{E}[\langle X_i\varepsilon_i, X_{i'}\varepsilon_{i'}\rangle] = \langle \mathsf{E}[X_i\varepsilon_i], \mathsf{E}[X_{i'}\varepsilon_{i'}]\rangle = 0$. For the last sum, one can see that $\mathsf{E}[\langle X_i\varepsilon_{i'}, X_{i_0}\varepsilon_{i_0'}\rangle] = \mathsf{E}[\|X\|^2]\mathsf{E}[\varepsilon^2] < \infty$ if either $(i,i') = (i_0, i_0')$ or $(i,i') = (i_0', i_0)$ and $\mathsf{E}[\langle X_i\varepsilon_{i'}, X_{i_0}\varepsilon_{i_0'}\rangle] = 0$ otherwise. This implies that $\mathsf{E}[\|\bar{X}\bar{\varepsilon}\|^2] \le Cn^{-2}$. □

**Lemma 19.** *We have that* $\mathsf{E}[\langle \bar{X}, \Gamma_{h_n}^{-1}X_0\rangle^2] = n^{-1}h_n$.

*Proof.* It follows from

$$\mathsf{E}[\langle \bar{X}, \Gamma_{h_n}^{-1}X_0\rangle^2] = n^{-1}\mathsf{E}[\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2] = n^{-1}\mathsf{E}[\langle X_i^{\otimes 2}\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0]$$
$$= n^{-1}\mathsf{E}[\mathsf{E}^{X_0}[\langle X_i^{\otimes 2}\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle]] = n^{-1}\mathsf{E}[\langle \Gamma\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle]$$
$$= n^{-1}\mathsf{E}[\langle \Gamma_{h_n}^{-1}X_0, X_0\rangle] = n^{-1}\mathsf{E}[t_{h_n}(X_0)]$$
$$= h_n/n.$$

□

### 3.9.2 Concistency of the functional principal component regression (FPCR) estimator

The asymptotics of the FPCR estimator $\hat{\beta}_{h_n}$ of $\beta$ is based on the following decomposition:

$$\hat{\beta}_{h_n} - \beta = (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1})U_n + \Gamma_{h_n}^{-1}U_n + (\hat{\Pi}_{h_n} - \Pi_{h_n})\beta + \Pi_{h_n}\beta - \beta. \tag{3.38}$$

In this subsection, we suppose that Conditions (A1)-(A5) in the main text hold. The following lemmas are generalized results of Lemmas S3-S5 described in the supplement of [53] to the heteroscedastic models. Lemmas 20-21 can be proved in the same way while Lemma 22 needs a little more effort due to the centering.

**Lemma 20.** *As $n \to \infty$, we have the following:*

1. $\|\hat{\Pi}_{h_n} - \Pi_{h_n}\|_\infty = O_{\mathsf{P}}\left(n^{-1/2} \sum_{j=1}^{h_n} j \log j\right)$.

2. $\|\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}\|_\infty = O_{\mathsf{P}}\left(n^{-1/2} \sum_{j=1}^{h_n} \delta_j^{-1} j \log j\right)$.

3. (a) *Conditional on $X_0$.*

   *If $n^{-1/2} \sum_{j=1}^{h_n} \delta_j^{1/2} (j \log j)^{3/2} \to 0$, then for each $\eta > 0$,*

$$\mathsf{P}(\|(\hat{\Pi}_{h_n} - \Pi_{h_n})X_0\| > \eta | X_0) \xrightarrow{\mathsf{P}} 0.$$

   (b) *Unconditional on $X_0$.*

   $\|(\hat{\Pi}_{h_n} - \Pi_{h_n})X_0\| = O_{\mathsf{P}}\left(n^{-1/2} \sum_{j=1}^{h_n} \delta_j^{1/2} (j \log j)^{3/2}\right)$.

4. (a) *Conditional on $X_0$.*

   *If $n^{-1/2} \sum_{j=1}^{h_n} \delta_j^{-1/2} (j \log j)^{3/2} \to 0$, then for each $\eta > 0$,*

$$\mathsf{P}(\|(\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1})X_0\| > \eta | X_0) \xrightarrow{\mathsf{P}} 0.$$

   (b) *Unconditional on $X_0$.*

   $\|(\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1})X_0\| = O_{\mathsf{P}}\left(n^{-1/2} \sum_{j=1}^{h_n} \delta_j^{-1/2} (j \log j)^{3/2}\right)$.

**Lemma 21.** *We have $\|(\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1})U_n\| = O_{\mathsf{P}}\left(n^{-1} \sum_{j=1}^{h_n} \delta_j^{-1} j \log j\right)$.*

**Lemma 22.** *In general, we have*

$$\mathsf{E}[\|\Gamma_{h_n}^{-1} U_n\|^2] \leq C \left( n^{-1} \sum_{j=1}^{h_n} \gamma_j^{-2} \|\Lambda^{1/2} \phi_j\|^2 + n^{-2} \sum_{j=1}^{h_n} \gamma_j^{-1} \right).$$

*Furthermore, if $\sup_{j \in \mathbb{N}} \gamma_j^{-1} \|\Lambda^{1/2} \phi_j\|^2 < \infty$, we have that $\mathsf{E}[\|\Gamma_{h_n}^{-1} U_n\|^2] \leq C n^{-1} \sum_{j=1}^{h_n} \gamma_j^{-1}$.*

*Proof.* Notice that

$$\begin{aligned} \|\Gamma_{h_n}^{-1} U_n\|^2 &\leq 2\|\Gamma_{h_n}^{-1} \overline{X\varepsilon}\|^2 + 2\|\Gamma_{h_n}^{-1} \bar{X}\bar{\varepsilon}\|^2 \\ &= 2 \sum_{j=1}^{h_n} \gamma_j^{-2} \langle \overline{X\varepsilon}, \phi_j \rangle^2 + 2 \sum_{j=1}^{h_n} \gamma_j^{-2} \langle \bar{X}\bar{\varepsilon}, \phi_j \rangle^2. \end{aligned} \tag{3.39}$$

Let $\mathcal{X}_n \equiv \{X_i\}_{i=1}^n$ denote the observed regressors.

For the first term in (3.39), we compute its conditional expectation as

$$\begin{aligned}\mathsf{E}[\langle\overline{X\varepsilon},\phi_j\rangle^2|\mathcal{X}_n] &= n^{-2}\sum_{i=1}^n \mathsf{E}[\langle X_i\varepsilon_i,\phi_j\rangle^2|\mathcal{X}_n] + n^{-2}\sum_{i\neq i'}\mathsf{E}[\langle X_i\varepsilon_i,\phi_j\rangle\langle X_{i'}\varepsilon_{i'},\phi_j\rangle|\mathcal{X}_n]\\ &= n^{-1}\mathsf{E}[\langle X\varepsilon,\phi_j\rangle^2|\mathcal{X}_n]\end{aligned}$$

since

$$\begin{aligned}\mathsf{E}[\langle X_i\varepsilon_i,\phi_j\rangle\langle X_{i'}\varepsilon_{i'},\phi_j\rangle|\mathcal{X}_n] &= \langle X_i,\phi_j\rangle\langle X_{i'},\phi_j\rangle\mathsf{E}[\varepsilon_i\varepsilon_{i'}|\mathcal{X}_n]\\ &= \langle X_i,\phi_j\rangle\langle X_{i'},\phi_j\rangle\mathsf{E}[\varepsilon_i|X_i]\mathsf{E}[\varepsilon_{i'}|X_{i'}]\\ &= 0.\end{aligned}$$

This implies that $\mathsf{E}[\|\Gamma_{h_n}^{-1}\overline{X\varepsilon}\|^2|\mathcal{X}_n] = n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-2}\mathsf{E}[\langle X_1\varepsilon_1,\phi_j\rangle^2|\mathcal{X}_n]$. Then, from the fact that

$$\mathsf{E}[\langle X\varepsilon,\phi_j\rangle^2] = \mathsf{E}[\langle (X\varepsilon)^{\otimes 2}\phi_j,\phi_j\rangle] = \langle\Lambda\phi_j,\phi_j\rangle = \|\Lambda^{1/2}\phi_j\|^2 \le \|\Lambda\|_\infty,$$

its general second moment bound is obtained as

$$\mathsf{E}[\|\Gamma_{h_n}^{-1}\overline{X\varepsilon}\|^2] = n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-2}\|\Lambda^{1/2}\phi_j\|^2 \le \frac{\|\Lambda\|_\infty}{n}\sum_{j=1}^{h_n}\gamma_j^{-2}.$$

For the second term in (3.39), note that

$$\begin{aligned}\langle\bar{X}\bar{\varepsilon},\phi_j\rangle^2 &= n^{-4}\left(\sum_{i=1}^n\langle X_i\varepsilon_i,\phi_j\rangle + \sum_{i\neq i'}\langle X_i\varepsilon_{i'},\phi_j\rangle\right)^2\\ &\le 2n^{-4}\left(\sum_{i=1}^n\langle X_i\varepsilon_i,\phi_j\rangle\right)^2 + 2n^{-4}\left(\sum_{i\neq i'}\langle X_i\varepsilon_{i'},\phi_j\rangle\right)^2\\ &= 2n^{-4}\sum_{i=1}^n\langle X_i\varepsilon_i,\phi_j\rangle^2 + 2n^{-4}\sum_{i\neq i'}\langle X_i\varepsilon_i,\phi_j\rangle\langle X_{i'}\varepsilon_{i'},\phi_j\rangle\\ &\quad + 2n^{-4}\sum_{i\neq i',i_0\neq i_0'}\langle X_i\varepsilon_{i'},\phi_j\rangle\langle X_{i_0}\varepsilon_{i_0'},\phi_j\rangle.\end{aligned}$$

As above, we have $\mathsf{E}[\langle X_i\varepsilon_i,\phi_j\rangle^2] = \|\Lambda^{1/2}\phi_j\|^2$ and $\mathsf{E}[\langle X_i\varepsilon_i,\phi_j\rangle\langle X_{i'}\varepsilon_{i'},\phi_j\rangle] = 0$. For the last term in the previous display, if either $(i,i') = (i_0,i_0')$ or $(i,i') = (i_0',i_0)$, we have

$$\begin{aligned}\mathsf{E}[\langle X_i\varepsilon_{i'},\phi_j\rangle\langle X_{i_0}\varepsilon_{i_0'},\phi_j\rangle] &= \mathsf{E}[\langle X_i,\phi_j\rangle^2\varepsilon_{i'}^2] = \mathsf{E}[\langle X,\phi_j\rangle^2]\mathsf{E}[\varepsilon^2]\\ &= \gamma_j\mathsf{E}[\varepsilon^2]\end{aligned}$$

and $\mathsf{E}[\langle X_i\varepsilon_{i'}, \phi_j\rangle\langle X_{i_0}\varepsilon_{i_0'}, \phi_j\rangle] = 0$ otherwise. This implies that

$$\mathsf{E}[\|\Gamma_{h_n}^{-1}\bar{X}\bar{\varepsilon}\|^2] \le C\left(n^{-3}\sum_{j=1}^{h_n}\gamma_j^{-2}\|\Lambda^{1/2}\phi_j\|^2 + n^{-2}\sum_{j=1}^{h_n}\gamma_j^{-1}\right).$$

Combinig these two results, we have that

$$\mathsf{E}[\|\Gamma_{h_n}^{-1}U_n\|^2] \le C\left(n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-2}\|\Lambda^{1/2}\phi_j\|^2 + n^{-2}\sum_{j=1}^{h_n}\gamma_j^{-1}\right).$$

□

We state the consistency of the FPCR estimator $\hat{\beta}_{h_n}$ for the slope function $\beta$ in the following theorem, which can be proved in the same way as Theorem S1 in the supplement of [53]

**Theorem 9** (Consistency of the FPCR estimator)**.** *Suppose that* $h_n^{-1} + n^{-1/2}h_n^2\log h_n \to 0$ *as* $n \to \infty$*. Then, the FPCR estimator* $\hat{\beta}_{h_n}$ *converges to the slope function* $\beta$ *in probability in the sense that* $\|\hat{\beta}_{h_n} - \beta\| \xrightarrow{\mathsf{P}} 0$ *as* $n \to \infty$*.*

*Proof.* First, note that the remainder term related to $\mathcal{E}_j^c$ and $\mathcal{A}_{h_n}^c$ are negligible by Lemma 17. Then, by the decomposition (3.38), and Lemmas 20-22, we have that

$$\begin{aligned}
\|\hat{\beta}_{h_n} - \beta\| &\le \|(\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1})U_n\| + \|\Gamma_{h_n}^{-1}U_n\| + \|(\hat{\Pi}_{h_n} - \Pi_{h_n})\|_\infty\|\beta\| + \|\Pi_{h_n}\beta - \beta\| \\
&= O_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-1}j\log j\right) + O_{\mathsf{P}}\left(\left(n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-2}\right)^{1/2}\right) \\
&\quad + O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n}j\log j\right) + O\left(\left(\sum_{j>h_n}\langle\beta,\phi_j\rangle^2\right)^{1/2}\right)
\end{aligned}$$

We note the following convergences, which can be derived from some algebra. First, as $n \to \infty$, $n^{-1}\sum_{j=1}^{h_n}(j\log j)^2 \le n^{-1}h_n^3(\log h_n)^2 \le (n^{-1/2}h_n^2\log h_n)^2 \to 0$, we have

$$n^{-1}\sum_{j=1}^{h_n}\delta_j^{-1}j\log j \le \left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}\right)^{1/2}\left\{n^{-1}\sum_{j=1}^{h_n}(j\log j)^2\right\} \to 0$$

by Cauchy-Schwarz inequality. For the rest of terms in the above display, we note that $n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-2} \le n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2} \to 0$, $n^{-1/2}\sum_{j=1}^{h_n}j\log j \le n^{-1/2}h_n^2\log h_n \to 0$, and $\sum_{j>h_n}\langle\beta,\phi_j\rangle^2 \to 0$, as $n \to \infty$. We therefore have that $\mathsf{P}(\|\hat{\beta}_{h_n} - \beta\| > \eta) \to 0$ as $n \to \infty$. □

### 3.9.3 Central limit theorem

The proof of the CLT is again based on the decomposition (3.38). In this subsection, we suppose that Conditions (A1)-(A7) in the main paper hold. Upon scaling $\sqrt{n/s_{h_n}(X_0)}$, Section 3.9.3.1 describes the convergences of bias terms $(\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1})U_n$, $(\hat{\Pi}_{h_n} - \Pi_{h_n})\beta$, $\Pi_{h_n}\beta - \beta$ in (3.38) to zero, while the proof for the weak convergence of the variance term $\Gamma_{h_n}^{-1}U_n$ in (3.38) is provided in Section 3.9.3.2. The consistency of the sample scaling $\hat{s}_{h_n}(X_0)$ is finally proved in Section 3.9.3.3.

#### 3.9.3.1 Bias terms

The below propositions 17-19 can be proved in the same way as [12, 53], and hence, the proofs are omitted here.

**Proposition 17.** *As $n \to \infty$, if $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}(j\log j)^2 \to 0$, we have that*

$$\mathsf{P}\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle(\hat{\Pi}_{h_n} - \Pi_{h_n})\beta, X_0\rangle| > \eta \middle| X_0\right) \xrightarrow{\mathsf{P}} 0.$$

**Proposition 18.** *As $n \to \infty$, suppose $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2} \to 0$. We then have that*

$$\mathsf{P}\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle(\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1})U_n, X_0\rangle| > \eta \middle| X_0\right) \xrightarrow{\mathsf{P}} 0.$$

**Proposition 19.** *For any sequence $\{\zeta_h\}_{h\in\mathbb{N}}$ such that $\{h^{-1}\zeta_h\}_{h\in\mathbb{N}}$ is non-decreasing, we have the following moment inequality:*

$$\frac{n}{h_n}\mathsf{E}[\langle\Pi_{h_n}\beta - \beta, X_0\rangle^2] \le \frac{n}{\zeta_{h_n}}\left(\sum_{j>h_n}\gamma_j\right)\sup_{j\in\mathbb{N}}\left(j^{-1}\zeta_j\langle\beta, \phi_j\rangle^2\right).$$

*Hence, if $\sup_{j\in\mathbb{N}}\left(j^{-1}\zeta_j\beta_j^2\right) < \infty$ and $n = O(\zeta_{h_n})$ as $n \to \infty$, then we have that*

$$\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\Pi_{h_n}\beta - \beta, X_0\rangle \xrightarrow{\mathsf{P}} 0.$$

#### 3.9.3.2 Weakly convergent term

We provide the proof for the weak convergence of the variance term as described in Proposition 15 of the main paper.

*Proof of Proposition 15 in the main paper.* Let $\mathsf{P}^{X_0} \equiv P(\cdot|X_0)$ and $\mathsf{E}^{X_0} \equiv \mathsf{E}[\cdot|X_0]$ respectively denote the conditional probability and expectation given the new predictor $X_0$. We first write $U_n = \tilde{U}_n - \bar{X}\bar{\varepsilon}$ where $\tilde{U}_n \equiv n^{-1}\sum_{i=1}^n X_i\varepsilon_i$. The contribution from the second term $\bar{X}\bar{\varepsilon}$ can be shown as negligible as follows. Note that

$$\langle \Gamma_{h_n}^{-1}\bar{X}\bar{\varepsilon}, X_0\rangle^2 \le n^{-4}2\left\langle \sum_{i=1}^n X_i\varepsilon_i, \Gamma_{h_n}^{-1}X_0\right\rangle^2 + n^{-4}2\left\langle \sum_{i\neq i'} X_i\varepsilon_{i'}, \Gamma_{h_n}^{-1}X_0\right\rangle^2.$$

Due to the independence of the sample $\{(X_i, Y_i)\}_{i=1}^n$ and $\mathsf{E}[X\varepsilon] = 0$, we have that

$$\begin{aligned}
n^{-1}\mathsf{E}^{X_0}\left[\left\langle \sum_{i=1}^n X_i\varepsilon_i, \Gamma_{h_n}^{-1}X_0\right\rangle^2\right] &= \mathsf{E}^{X_0}[\langle X_i\varepsilon_i, \Gamma_{h_n}^{-1}X_0\rangle^2] \\
&= \mathsf{E}^{X_0}[\langle (X_i\varepsilon_i)^{\otimes 2}\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle^2] \\
&= \langle \Lambda\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle = s_{h_n}(X_0);
\end{aligned}$$

similarly, since $E[X] = E[X\varepsilon] = 0$, the expected value of cross terms is given by

$$\begin{aligned}
\mathsf{E}^{X_0}\left[\left\langle \sum_{i\neq i'} X_i\varepsilon_{i'}, \Gamma_{h_n}^{-1}X_0\right\rangle^2\right] &= (n^2-n)\mathsf{E}^{X_0}[\langle X_i\varepsilon_{i'}, \Gamma_{h_n}^{-1}X_0\rangle^2] \\
&= (n^2-n)\mathsf{E}^{X_0}[\langle \varepsilon_{i'}^2 X_i^{\otimes 2}\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle] \\
&= (n^2-n)\mathsf{E}[\varepsilon^2]\langle \Gamma\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle \\
&\equiv (n^2-n)\mathsf{E}[\varepsilon^2]t_{h_n}(X_0).
\end{aligned}$$

These moments imply that

$$\begin{aligned}
&\mathsf{E}^{X_0}\left[\frac{n}{s_{h_n}(X_0)}\langle \Gamma_{h_n}^{-1}\bar{X}\bar{\varepsilon}, X_0\rangle^2\right] \\
\le& C\frac{n}{s_{h_n}(X_0)}\frac{s_{h_n}(X_0)}{n^3} + C\frac{n}{s_{h_n}(X_0)}\frac{(n^2-n)\mathsf{E}[\varepsilon^2]t_{h_n}(X_0)}{n^4} \\
\le& Cn^{-2} + Cn^{-1}\frac{t_{h_n}(X_0)}{s_{h_n}(X_0)} = O_{\mathsf{P}}(n^{-1})
\end{aligned}$$

due to the fact that $\mathsf{E}[t_{h_n}(X_0)] = h_n$ and Condition (A6).

We next consider the contribution from the term $\tilde{U}_n \equiv n^{-1}\sum_{i=1}^n X_i\varepsilon_i$. Write $Z_{i,n} = \langle X_i\varepsilon_i, \Gamma_{h_n}^{-1}X_0\rangle$ for $i = 1, \ldots, n$ so that

$$\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle \Gamma_{h_n}^{-1}\tilde{U}_n, X_0\rangle = (ns_{h_n}(X_0))^{-1/2}\sum_{i=1}^n Z_{i,n}.$$

Notice that $\mathsf{E}^{X_0}[Z_{i,n}] = \langle \mathsf{E}[X_i\varepsilon_i], \Gamma_{h_n}^{-1}X_0\rangle = 0$, and set $v_n^2 \equiv \sum_{i=1}^n \mathsf{E}^{X_0}[Z_{i,n}^2]$. To verify a Lindeberg condition (conditional on $X_0$), define a quantity $\mathcal{L}_n \equiv v_n^{-2}\sum_{i=1}^n \mathsf{E}^{X_0}[Z_{i,n}^2\mathbb{I}(|Z_{i,n}| > \tau v_n)]$ for $\tau > 0$. Proposition 15 will then follow by showing that $\mathcal{L}_n \xrightarrow{\mathsf{P}} 0$ holds as $n \to \infty$. For this purpose, it suffices to establish (3.40)-(3.42) below:

$$n^{-1}v_n^2 = s_{h_n}(X_0), \tag{3.40}$$

$$\mathsf{E}^{X_0}\left[\left(v_n^{-1}\max_{1\le i\le n}|Z_{i,n}|\right)^4\right] \xrightarrow{\mathsf{P}} 0 \quad \text{as} \quad n\to\infty, \tag{3.41}$$

$$\mathsf{E}^{X_0}\left[\left|\frac{n^{-1}\sum_{i=1}^n Z_{i,n}^2}{s_{h_n}(X_0)} - 1\right|^2\right] \xrightarrow{\mathsf{P}} 0 \quad \text{as} \quad n\to\infty. \tag{3.42}$$

By (3.40) and (3.42), it also holds that

$$\mathsf{E}^{X_0}\left[\left|v_n^{-2}\sum_{i=1}^n Z_{i,n}^2 - 1\right|^2\right] \xrightarrow{\mathsf{P}} 0$$

as $n\to\infty$. For notational convenience, we write $A_n = v_n^{-2}\sum_{i=1}^n Z_{i,n}^2$, $B_n = v_n^{-2}\sum_{i=1}^n Z_{i,n}^2\mathbb{I}(|Z_{i,n}| > \tau v_n)$, and $C_n = v_n^{-1}\max_{1\le i\le n}|Z_{i,n}|$. Due to a subsequence argument (cf. [4], Theorem 20.5), the convergences in probability in (3.41)-(3.42) can be treated as almost sure convergence along a subsequence. Along any such subsequence, it then holds that $B_n \to 0$ in probability (with respect to $\mathsf{P}^{X_0}$) by $B_n \le A_n\mathbb{I}(C_n > \tau) \to 1\cdot 0 = 0$ and it also holds that $\{B_n\}$ is uniformly integrable (with respect to $\mathsf{P}^{X_0}$) due to $B_n \le A_n$ and the convergence from (3.42). Along the subsequence, on an almost sure set, we have that $\mathcal{L}_n = \mathsf{E}^{X_0}[B_n] \to \mathsf{E}^{X_0}[0] = 0$, which verifies the Lindeberg condition. That is, it follows that

$$\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle \Gamma_{h_n}^{-1}\tilde{U}_n, X_0\rangle = \{ns_{h_n}(X_0)\}^{-1/2}\sum_{i=1}^n Z_{i,n} = v_n^{-1}\sum_{i=1}^n Z_{i,n} \xrightarrow{\mathsf{d}} \mathsf{N}(0,1),$$

with respect to $\mathsf{P}^{X_0}$ by (3.40). As the CDF $\Phi$ of the standard normal distribution is continuous, by Polya's theorem (cf. [2], Theorem 9.1.4), we have

$$\sup_{y\in\mathbb{R}}\left|\mathsf{P}^{X_0}\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle \Gamma_{h_n}^{-1}\tilde{U}_n, X_0\rangle \le y\right) - \Phi(y)\right| \xrightarrow{\mathsf{P}} 0,$$

establishing Proposition 15.

We next show that (3.40)-(3.42) hold. The equation (3.40) follows from $\mathsf{E}^{X_0}[Z_{i,n}^2] = \langle \Lambda\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle = s_{h_n}(X_0)$. To prove the convergence in (3.41), we note that

$$|Z_{i,n}| = |\langle \Lambda_{h_n}^{-1/2}X_i\varepsilon_i, \Lambda_{h_n}^{1/2}\Gamma_{h_n}^{-1}X_0\rangle| \leq \|\Lambda_{h_n}^{-1/2}X_i\varepsilon_i\|\|\Lambda_{h_n}^{1/2}\Gamma_{h_n}^{-1}X_0\|,$$

which implies that

$$\begin{aligned} v_n^{-1}\max_{1\leq i\leq n}|Z_{i,n}| &\leq n^{-1/2}\max_{1\leq i\leq n}\|\Lambda_{h_n}^{-1/2}X_i\varepsilon_i\|\frac{\|\Lambda_{h_n}^{1/2}\Gamma_{h_n}^{-1}X_0\|}{(n^{-1}v_n^2)^{1/2}} \\ &\leq n^{-1/2}\max_{1\leq i\leq n}\|\Lambda_{h_n}^{-1/2}X_i\varepsilon_i\|. \end{aligned}$$

since $\|\Lambda_{h_n}^{1/2}\Gamma_{h_n}^{-1}X_0\| \leq \|\Lambda^{1/2}\Gamma_{h_n}^{-1}X_0\|$. By Jensen's inequality, it then holds that

$$\begin{aligned} \|\Lambda_{h_n}^{-1/2}X_i\varepsilon_i\|^2 = \langle \Lambda_{h_n}^{-1}X_i\varepsilon_i, X_i\varepsilon_i\rangle &= \sum_{j=1}^{h_n}\lambda_j^{-1}\langle X_i\varepsilon_i, \psi_j\rangle^2 \\ &\leq \sqrt{h_n\sum_{j=1}^{h_n}\lambda_j^{-2}\langle X_i\varepsilon_i, \psi_j\rangle^4}. \end{aligned}$$

By the finite fourth moment assumption (A7) on $X\varepsilon$, we see that

$$\mathsf{E}\left[\max_{1\leq i\leq n}\|\Lambda_{h_n}^{-1/2}X_i\varepsilon_i\|^4\right] \leq \mathsf{E}\left[h_n\sum_{i=1}^{n}\sum_{j=1}^{h_n}\lambda_j^{-2}\langle X_i\varepsilon_i, \psi_j\rangle^4\right] \leq Cnh_n^2,$$

which implies that

$$\mathsf{E}^{X_0}\left[\left(v_n^{-1}\max_{1\leq i\leq n}|Z_{i,n}|\right)^4\right] \leq n^{-2}\mathsf{E}\left[\max_{1\leq i\leq n}\|\Lambda_{h_n}^{-1/2}X_i\varepsilon_i\|^4\right] \leq Cn^{-1}h_n^2 \to 0$$

as $n\to\infty$.

Finally, we verify the convergence in (3.42). We note that

$$n^{-1}\sum_{i=1}^{n}Z_{i,n}^2 = \langle \tilde{\Lambda}_n\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle = \langle (\tilde{\Lambda}_n-\Lambda)\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle + s_{h_n}(X_0),$$

which implies that

$$\begin{aligned} \mathsf{E}^{X_0}\left[\left|\frac{n^{-1}\sum_{i=1}^{n}Z_{i,n}^2}{s_{h_n}(X_0)}-1\right|^2\right] &= \mathsf{E}^{X_0}\left[\left|s_{h_n}(X_0)^{-1}\langle(\tilde{\Lambda}_n-\Lambda)\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle\right|^2\right] \\ &\leq \{h_n s_{h_n}(X_0)^{-1}\}^2\mathsf{E}^{X_0}[\|\tilde{\Lambda}_n-\Lambda\|_\infty^2](h_n^{-2}\|\Gamma_{h_n}^{-1}X_0\|^4) \\ &= O_{\mathsf{P}}\left(\left(n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^2\right) \end{aligned}$$

since $\mathsf{E}[\|\tilde{\Lambda}_n - \Lambda\|_\infty^2] = O(n^{-1})$ (cf. [32], Theorem 2.5) from the finite fourth moment assumption (A7) and $\mathsf{E}[\|\Gamma_{h_n}^{-1} X_0\|^2] = \sum_{j=1}^{h_n} \gamma_j^{-2} \mathsf{E}[\langle X_0, \phi_j\rangle^2] = \sum_{j=1}^{h_n} \gamma_j^{-1}$. One can find that $n^{-1/2} h_n^{-1} \sum_{j=1}^{h_n} \gamma_j^{-1}$ is dominated by $h_n^{-1/2} \left(n^{-1} \sum_{j=1}^{h_n} \delta_j^{-2}\right)^{1/2}$ by applying Cauchy-Schwarz inequality. Hence, due to (A5), as $n \to \infty$, we have that

$$\mathsf{E}^{X_0}\left[\left|\frac{n^{-1}\sum_{i=1}^n Z_{i,n}^2}{s_{h_n}(X_0)} - 1\right|^2\right] \xrightarrow{\mathsf{P}} 0.$$

□

#### 3.9.3.3 Scaling term

We investigate the consistency of the ratio of $\hat{s}_{h_n}(X_0)$ over $s_{h_n}(X_0)$ to 1 either conditionally or unconditionally on $X_0$. To obtain this consistency, two approximations of $\Lambda$ and $\Gamma^{-1}$ appear, which complicates the proofs. The proof is based on the following decomposition of the empirical scaling $\hat{s}_{h_n}(X_0)$:

$$\begin{aligned}
\hat{s}_{h_n}(X_0) =& \langle \hat{\Lambda}_{n,k_n} \hat{\Gamma}_{h_n}^{-1} X_0, \hat{\Gamma}_{h_n}^{-1} X_0\rangle \\
=& \langle \hat{\Lambda}_{n,k_n} (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0, \hat{\Gamma}_{h_n}^{-1} X_0\rangle + \langle \hat{\Lambda}_{n,k_n} \Gamma_{h_n}^{-1} X_0, \hat{\Gamma}_{h_n}^{-1} X_0\rangle \\
=& \langle \hat{\Lambda}_{n,k_n} (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0, (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0\rangle + \langle \hat{\Lambda}_{n,k_n} (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0, \Gamma_{h_n}^{-1} X_0\rangle \\
&+ \langle \hat{\Lambda}_{n,k_n} \Gamma_{h_n}^{-1} X_0, (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0\rangle + \langle \hat{\Lambda}_{n,k_n} \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0\rangle \\
=& \langle \hat{\Lambda}_{n,k_n} (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0, (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0\rangle \\
&+ 2\langle \hat{\Lambda}_{n,k_n} \Gamma_{h_n}^{-1} X_0, (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0\rangle \\
&+ \langle \hat{\Lambda}_{n,k_n} \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0\rangle \\
=& \langle (\hat{\Lambda}_{n,k_n} - \Lambda)(\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0, (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0\rangle \qquad (3.43) \\
&+ \langle \Lambda(\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0, (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0\rangle \\
&+ 2\langle (\hat{\Lambda}_{n,k_n} - \Lambda) \Gamma_{h_n}^{-1} X_0, (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0\rangle \\
&+ 2\langle \Lambda \Gamma_{h_n}^{-1} X_0, (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0\rangle \\
&+ \langle (\hat{\Lambda}_{n,k_n} - \Lambda) \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0\rangle + \langle \Lambda \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0\rangle.
\end{aligned}$$

We see by writing $\hat{\varepsilon}_{i,k_n} = \varepsilon_i - \langle \hat{\beta}_{k_n} - \beta, X_i \rangle$ that

$$\begin{aligned}\hat{\Lambda}_{n,k_n} - \Lambda &= n^{-1}\sum_{i=1}^{n}(X_i\hat{\varepsilon}_{i,k_n})^{\otimes 2} - (\overline{X\hat{\varepsilon}})_{n,k_n}^{\otimes 2} - \Lambda \\ &= \tilde{\Lambda}_n - \Lambda + n^{-1}\sum_{i=1}^{n} X_i^{\otimes 2}\langle \hat{\beta}_{k_n} - \beta, X_i\rangle^2 \\ &\quad - 2n^{-1}\sum_{i=1}^{n}\{(X_i\varepsilon_i)\otimes X_i\}\langle \hat{\beta}_{k_n} - \beta, X_i\rangle - (\overline{X\hat{\varepsilon}})_{n,k_n}^{\otimes 2}\end{aligned} \tag{3.44}$$

where $\tilde{\Lambda}_n \equiv n^{-1}\sum_{i=1}^{n}(X_i\varepsilon_i)^{\otimes 2}$ and

$$\begin{aligned}(\overline{X\hat{\varepsilon}})_{n,k_n} &\equiv n^{-1}\sum_{i=1}^{n} X_i\hat{\varepsilon}_{i,k_n} = n^{-1}\sum_{i=1}^{n} X_i\varepsilon_i - n^{-1}\sum_{i=1}^{n}\langle \hat{\beta}_{k_n} - \beta, X_i\rangle X_i \\ &= \overline{X\varepsilon} - \tilde{\Gamma}_n(\hat{\beta}_{k_n} - \beta)\end{aligned} \tag{3.45}$$

with $\overline{X\varepsilon} \equiv n^{-1}\sum_{i=1}^{n} X_i\varepsilon_i$ and $\tilde{\Gamma}_n \equiv n^{-1}\sum_{i=1}^{n} X_i^{\otimes 2}$.

In Lemmas 23-25, we study the convergence rates of each term in (3.43) by using the decomposition (3.44) of $\hat{\Lambda}_{n,k_n} - \Lambda$ In the following lemmas, we suppose that $\|\hat{\beta}_{k_n} - \beta\| \xrightarrow{\mathsf{P}} 0$ as $n \to \infty$.

**Lemma 23.**

1. *We have that* $(\overline{X\hat{\varepsilon}})_{n,k_n} = o_{\mathsf{P}}(1) + O_{\mathsf{P}}(\|\hat{\beta}_{k_n} - \beta\|)$ *so that* $(\overline{X\hat{\varepsilon}})_{n,k_n} \xrightarrow{\mathsf{P}} 0$ *as* $n \to \infty$.

2. *As* $n \to \infty$, *if* $n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1} \to 0$ *(which is implied by Condition (A5)), we have that*

$$\mathsf{P}(s_{h_n}(X_0)^{-1}\langle(\overline{X\hat{\varepsilon}})_{n,k_n}, \Gamma_{h_n}^{-1}X_0\rangle^2 > \eta | X_0) \xrightarrow{\mathsf{P}} 0.$$

*Proof.* The first part follows since

$$\begin{aligned}\|(\overline{X\hat{\varepsilon}})_{n,k_n}\| &\le \|\overline{X\varepsilon}\| + \|\tilde{\Gamma}_n - \Gamma\|_\infty\|\hat{\beta}_{k_n} - \beta\| + \|\Gamma\|_\infty\|\hat{\beta}_{k_n} - \beta\| \\ &= o_{\mathsf{P}}(1) + O_{\mathsf{P}}(\|\hat{\beta}_{k_n} - \beta\|)\end{aligned}$$

by the law of large numbers (cf. [33], Theorem 7.2).

For the second part, note that $\mathsf{E}[\|\overline{X\varepsilon}\|^2] = O(n^{-1})$ from Theorem 2.3 of [32] under Condition (A7) and $\mathsf{E}[\|\Gamma_{h_n}^{-1}X_0\|^2] = \sum_{j=1}^{h_n}\gamma_j^{-1}$. This implies that

$\mathsf{E}[\langle\overline{X\varepsilon}, \Gamma_{h_n}^{-1}X_0\rangle^2] = O\left(n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)$, and hence, we have that

$$s_{h_n}(X_0)^{-1}\langle\overline{X\varepsilon}, \Gamma_{h_n}^{-1}X_0\rangle^2 = O_{\mathsf{P}}\left(n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right),$$

$$\mathsf{E}\left[s_{h_n}(X_0)^{-1}\langle\overline{X\varepsilon}, \Gamma_{h_n}^{-1}X_0\rangle^2\middle|X_0\right] = O_{\mathsf{P}}\left(n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right),$$

Next, we observe that

$$\langle\tilde{\Gamma}_n(\hat{\beta}_{k_n}-\beta), \Gamma_{h_n}^{-1}X_0\rangle = \langle(\hat{\beta}_{k_n}-\beta), \tilde{\Gamma}_n\Gamma_{h_n}^{-1}X_0\rangle$$

and $\tilde{\Gamma}_n\Gamma_{h_n}^{-1}X_0 = n^{-1}\sum_{i=1}^{n}\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle X_i$. By Cauchy-Schwarz inequality (for both arithmetic mean and expectation), we have that

$$\begin{aligned}\mathsf{E}[\|\tilde{\Gamma}_n\Gamma_{h_n}^{-1}X_0\|]^2 &\le \mathsf{E}\left[\left(n^{-1}\sum_{i=1}^{n}\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2\right)^{1/2}\left(n^{-1}\sum_{i=1}^{n}\|X_i\|^2\right)^{1/2}\right]^2\\ &\le \mathsf{E}\left[\left(n^{-1}\sum_{i=1}^{n}\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2\right)\right]\mathsf{E}\left[\left(n^{-1}\sum_{i=1}^{n}\|X_i\|^2\right)\right]\end{aligned}$$

and similarly, we have the conditional version as

$$\mathsf{E}[\|\tilde{\Gamma}_n\Gamma_{h_n}^{-1}X_0\||X_0]^2 \le \mathsf{E}\left[\left(n^{-1}\sum_{i=1}^{n}\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2\right)\middle|X_0\right]\mathsf{E}\left[\left(n^{-1}\sum_{i=1}^{n}\|X_i\|^2\right)\middle|X_0\right].$$

Since $\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle = \sum_{j=1}^{h_n}\gamma_j^{-1}\langle X_i, \phi_j\rangle\langle X_0, \phi_j\rangle$ and the FPC scores $\xi_j$ are uncorrelated random variables with mean zero and variance $\gamma_j$, we see that

$$\begin{aligned}\mathsf{E}\left[n^{-1}\sum_{i=1}^{n}\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2\right] &= n^{-1}\sum_{i=1}^{n}\mathsf{E}\left[\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2\right]\\ &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{h_n}\gamma_j^{-2}\mathsf{E}[\langle X_i, \phi_j\rangle^2]\mathsf{E}[\langle X_0, \phi_j\rangle^2]\\ &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{h_n}1 = h_n.\end{aligned}$$

Finally, since $\mathsf{E}\left[n^{-1}\sum_{i=1}^{n}\|X_i\|^2|X_0\right] = \mathsf{E}\left[n^{-1}\sum_{i=1}^{n}\|X_i\|^2\right] = \mathsf{E}\left[\|X\|^2\right] = \mathrm{tr}(\Gamma) < \infty$, we obtain

$$\mathsf{E}[s_{h_n}(X_0)^{-1/2}\|\tilde{\Gamma}_n\Gamma_{h_n}^{-1}X_0\||X_0] = O_{\mathsf{P}}(1).$$

Note that $\mathsf{P}(\|\hat{\beta}_{k_n} - \beta\| > \eta | X_0) = \mathsf{P}(\|\hat{\beta}_{k_n} - \beta\| > \eta) \to 0$ and

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathsf{P}(s_{h_n}(X_0)^{-1/2} \|\tilde{\Gamma}_n \Gamma_{h_n}^{-1} X_0\| > M | X_0) = 0.$$

This implies that

$$\mathsf{P}(s_{h_n}(X_0)^{-1/2} |\langle \tilde{\Gamma}_n(\hat{\beta}_{k_n} - \beta), \Gamma_{h_n}^{-1} X_0 \rangle| > \eta | X_0) \xrightarrow{\mathsf{P}} 0.$$

Finally, due to the decomposition (3.45), we have the desired result. □

**Lemma 24.**

1. *We have*

$$n^{-1} \sum_{i=1}^{n} X_i^{\otimes 2} \langle \hat{\beta}_{k_n} - \beta, X_i \rangle^2 = O_{\mathsf{P}}\left( \|\hat{\beta}_{k_n} - \beta\|^2 \right).$$

2. *For each $\eta > 0$, we have*

$$\begin{aligned} &\mathsf{P}\left( s_{h_n}(X_0)^{-1} \left| \left\langle \left( n^{-1} \sum_{i=1}^{n} X_i^{\otimes 2} \langle \hat{\beta}_{k_n} - \beta, X_i \rangle^2 \right) \Gamma_{h_n}^{-1} X_0, (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0 \right\rangle \right| > \eta \middle| X_0 \right) \\ =& o_{\mathsf{P}}\left( n^{-1/2} h_n^{-1/2} \sum_{j=1}^{h_n} \delta_j^{-1/2} (j \log j)^{3/2} + 1 \right). \end{aligned}$$

3. *For each $\eta > 0$, we have*

$$\mathsf{P}\left( s_{h_n}(X_0)^{-1} \left| \left\langle \left( n^{-1} \sum_{i=1}^{n} X_i^{\otimes 2} \langle \hat{\beta}_{k_n} - \beta, X_i \rangle^2 \right) \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \right\rangle \right| > \eta \middle| X_0 \right) \xrightarrow{\mathsf{P}} 0.$$

*Proof.* Note that the second term in the decomposition (3.44) can be bounded as

$$\left\| n^{-1} \sum_{i=1}^{n} X_i^{\otimes 2} \langle \hat{\beta}_{k_n} - \beta, X_i \rangle^2 \right\| \leq n^{-1} \sum_{i=1}^{n} \|X_i\|^4 \|\hat{\beta}_{k_n} - \beta\|^2 = O_{\mathsf{P}}(\|\hat{\beta}_{k_n} - \beta\|^2),$$

which proves the first part.

For the second part, note that

$$\begin{aligned}L_n \equiv & \left\langle \left( n^{-1} \sum_{i=1}^{n} X_i^{\otimes 2} \langle \hat{\beta}_{k_n} - \beta, X_i \rangle^2 \right) \Gamma_{h_n}^{-1} X_0, Q_n \right\rangle \\ = & n^{-1} \sum_{i=1}^{n} \langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle \langle X_i, Q_n \rangle \langle \hat{\beta}_{k_n} - \beta, X_i \rangle^2 \\ = & n^{-1} \sum_{i=1}^{n} \langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle \langle X_i, Q_n \rangle \langle X_i^{\otimes 2} (\hat{\beta}_{k_n} - \beta), \hat{\beta}_{k_n} - \beta \rangle \\ = & \left\langle n^{-1} \sum_{i=1}^{n} \langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle \langle X_i, Q_n \rangle X_i^{\otimes 2} (\hat{\beta}_{k_n} - \beta), \hat{\beta}_{k_n} - \beta \right\rangle\end{aligned}$$

where $Q_n \equiv (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0$. This implies that

$$|L_n| \leq \left( n^{-1} \sum_{i=1}^{n} \|X_i\|^3 |\langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle| \right) \|Q_n\| \|\hat{\beta}_{k_n} - \beta\|^2.$$

We note from Cauchy-Schwarz inequality that

$$\mathsf{E}[\|X_i\|^3 |\langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle|]^2 \leq \mathsf{E}[\|X_i\|^4] \mathsf{E}[\|X_i\|^2 \langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle^2].$$

Since the FPC scores $\xi_j$ are uncorrelated random variables with mean zero and variance 1, we have from the independence between $\mathcal{X}_n \equiv \{X_i\}_{i=1}^n$ and $X_0$ that

$$\mathsf{E}[\|X_i\|^2 \langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle^2] = \sum_{j=1}^{h_n} \gamma_j^{-1} \mathsf{E}[\|X_i\|^2 \langle X_i, \phi_j \rangle^2].$$

By Condition (A2) and Cauchy-Schwarz inequality, we see that

$$\mathsf{E}[\|X_i\|^2 \langle X_i, \phi_j \rangle^2] \leq \mathsf{E}[\|X_i\|^4]^{1/2} \mathsf{E}[\langle X_i \phi_j \rangle^4]^{1/2} \leq C \gamma_j,$$

which implies that $\mathsf{E}[\|X_i\|^3 |\langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle|] = O(h_n^{1/2})$. We then have

$$\mathsf{E}\left[ s_{h_n}(X_0)^{-1} \left( n^{-1} \sum_{i=1}^{n} \|X_i\|^3 |\langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle| \right) \Big| X_0 \right] = O_{\mathsf{P}}(h_n^{-1/2}),$$

under Condition (A6). Therefore, due to Lemma 20, we have the desired result.

Lastly, note that

$$
\begin{aligned}
L_n \equiv & \left\langle \left( n^{-1} \sum_{i=1}^{n} X_i^{\otimes 2} \langle \hat{\beta}_{k_n} - \beta, X_i \rangle^2 \right) \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \right\rangle \\
=& n^{-1} \sum_{i=1}^{n} \langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle^2 \langle \hat{\beta}_{k_n} - \beta, X_i \rangle^2 \\
=& n^{-1} \sum_{i=1}^{n} \langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle^2 \langle X_i^{\otimes 2} (\hat{\beta}_{k_n} - \beta), \hat{\beta}_{k_n} - \beta \rangle \\
=& \left\langle n^{-1} \sum_{i=1}^{n} \langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle^2 X_i^{\otimes 2} (\hat{\beta}_{k_n} - \beta), \hat{\beta}_{k_n} - \beta \right\rangle .
\end{aligned}
$$

This implies that

$$
|L_n| \leq \left( n^{-1} \sum_{i=1}^{n} \|X_i\|^2 \langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle^2 \right) \|\hat{\beta}_{k_n} - \beta\|^2 .
$$

Since the FPC scores $\xi_j$ are uncorrelated random variables with mean zero and variance 1, we have from the independence between $\mathcal{X}_n \equiv \{X_i\}_{i=1}^n$ and $X_0$ that

$$
\mathsf{E}[\|X_i\|^2 \langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle^2] = \sum_{j=1}^{h_n} \gamma_j^{-1} \mathsf{E}[\|X_i\|^2 \langle X_i, \phi_j \rangle^2].
$$

By Condition (A2) and Cauchy-Schwarz inequality, we see that

$$
\mathsf{E}[\|X_i\|^2 \langle X_i, \phi_j \rangle^2] \leq \mathsf{E}[\|X_i\|^4]^{1/2} \mathsf{E}[\langle X_i, \phi_j \rangle^4]^{1/2} \leq C\gamma_j,
$$

which implies that $\mathsf{E}[\|X_i\|^2 \langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle^2] = O(h_n)$. We then have

$$
\mathsf{E}\left[ s_{h_n}(X_0)^{-1} \left( n^{-1} \sum_{i=1}^{n} \|X_i\|^2 \langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle^2 \right) \middle| X_0 \right] = O_{\mathsf{P}}(1),
$$

under Condition (A6). Therefore, we have the desired result. □

**Lemma 25.**

1. *We have* $n^{-1} \sum_{i=1}^{n} \{(X_i \varepsilon_i) \otimes X_i\} \langle \hat{\beta}_{k_n} - \beta, X_i \rangle = O_{\mathsf{P}}(\|\hat{\beta}_{k_n} - \beta\|)$.

2. *For each* $\eta > 0$, *we have*

$$
\begin{aligned}
& \mathsf{P}\left( \left\langle \left( n^{-1} \sum_{i=1}^{n} \{(X_i \varepsilon_i) \otimes X_i\} \langle \hat{\beta}_{k_n} - \beta, X_i \rangle \right) \Gamma_{h_n}^{-1} X_0, (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) X_0 \right\rangle > \eta \middle| X_0 \right) \\
=& o_{\mathsf{P}}\left( n^{-1/2} h_n^{-1/2} \sum_{j=1}^{h_n} \delta_j^{-1/2} (j \log j)^{3/2} + 1 \right).
\end{aligned}
$$

*3. For each $\eta > 0$, we have*

$$\mathsf{P}\left(s_{h_n}(X_0)^{-1}\left|\left\langle\left(n^{-1}\sum_{i=1}^{n}\{(X_i\varepsilon_i)\otimes X_i\}\langle\hat{\beta}_{k_n}-\beta,X_i\rangle^2\right)\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\right\rangle\right|>\eta\middle|X_0\right)\xrightarrow{\mathsf{P}}0.$$

*Proof.* We first observe that

$$\begin{aligned}&\left\|n^{-1}\sum_{i=1}^{n}\{(X_i\varepsilon_i)\otimes X_i\}\langle\hat{\beta}_{k_n}-\beta,X_i\rangle\right\|\\ \le& n^{-1}\sum_{i=1}^{n}\|X_i\varepsilon\|\|X_i\|^2\|\hat{\beta}_{k_n}-\beta\|\\ \le&\left(n^{-1}\sum_{i=1}^{n}\|X_i\varepsilon\|^2\right)^{1/2}\left(n^{-1}\sum_{i=1}^{n}\|X_i\|^4\right)\|\hat{\beta}_{k_n}-\beta\|\\ =&O_{\mathsf{P}}(\|\hat{\beta}_{k_n}-\beta\|)\end{aligned}$$

since $\mathsf{E}[\|X\varepsilon\|^2]<\infty$ and $\mathsf{E}[\|X\|^4]<\infty$.

For the second part, note that

$$\begin{aligned}L_n&\equiv\left\langle\left(n^{-1}\sum_{i=1}^{n}\{(X_i\varepsilon_i)\otimes X_i\}\langle\hat{\beta}_{k_n}-\beta,X_i\rangle\right)\Gamma_{h_n}^{-1}X_0,Q_n\right\rangle\\ &=n^{-1}\sum_{i=1}^{n}\langle\hat{\beta}_{k_n}-\beta,X_i\varepsilon_i\rangle\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle\langle X_i,Q_n\rangle\\ &=\left\langle n^{-1}\sum_{i=1}^{n}\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle\langle X_i,Q_n\rangle X_i\varepsilon_i,\hat{\beta}_{k_n}-\beta\right\rangle\end{aligned}$$

where $Q_n\equiv(\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1})X_0$. This implies that

$$|L_n|\le\left(n^{-1}\sum_{i=1}^{n}|\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle|\|X_i\|\|X_i\varepsilon_i\|\right)\|Q_n\|\|\hat{\beta}_{k_n}-\beta\|$$

We note from Cauchy-Schwarz inequality that

$$\mathsf{E}[|\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle|\|X_i\|\|X_i\varepsilon_i\|]^2\le\mathsf{E}[\|X_i\|^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2]\mathsf{E}[\|X_i\varepsilon_i\|^2].$$

Since the FPC scores $\xi_j$ are uncorrelated random variables with mean zero and variance 1, we have from the independence between $\mathcal{X}_n\equiv\{X_i\}_{i=1}^n$ and $X_0$ that

$$\mathsf{E}[\|X_i\|^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2]=\sum_{j=1}^{h_n}\gamma_j^{-1}\mathsf{E}[\|X_i\|^2\langle X_i,\phi_j\rangle^2].$$

By Condition (A2) and Cauchy-Schwarz inequality, we see that

$$\mathsf{E}[\|X_i\|^2\langle X_i\phi_j\rangle^2] \le \mathsf{E}[\|X_i\|^4]^{1/2}\mathsf{E}[\langle X_i\phi_j\rangle^4]^{1/2} \le C\gamma_j,$$

which implies that $\mathsf{E}[|\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle|\|X_i\|\|X_i\varepsilon_i\|] = O(h_n^{1/2})$. We then have

$$\mathsf{E}\left[s_{h_n}(X_0)^{-1}\left(n^{-1}\sum_{i=1}^{n}|\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle|\|X_i\|\|X_i\varepsilon_i\|\right)\Big|X_0\right] = O_{\mathsf{P}}(h_n^{-1/2}),$$

under Condition (A6). Therefore, due to Lemma 20, we have the desired result.

Lastly, note that

$$\begin{aligned}L_n \equiv &\left\langle\left(n^{-1}\sum_{i=1}^{n}\{(X_i\varepsilon_i)\otimes X_i\}\langle\hat{\beta}_{k_n}-\beta, X_i\rangle\right)\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\right\rangle\\ =&n^{-1}\sum_{i=1}^{n}\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2\langle\hat{\beta}_{k_n}-\beta, X_i\varepsilon_i\rangle.\end{aligned}$$

This implies that

$$|L_n| \le \left(n^{-1}\sum_{i=1}^{n}\|X_i\varepsilon_i\|\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2\right)\|\hat{\beta}_{k_n}-\beta\|.$$

Since the FPC scores $\xi_j$ are uncorrelated random variables with mean zero and variance 1, we have from the independence between $\{(X_i, \varepsilon_i)\}_{i=1}^n$ and $X_0$ that

$$\mathsf{E}[\|X_i\varepsilon_i\|\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2] = \sum_{j=1}^{h_n}\gamma_j^{-1}\mathsf{E}[\|X_i\varepsilon_i\|\langle X_i\phi_j\rangle^2].$$

By Condition (A2) and Cauchy-Schwarz inequality, we see that

$$\mathsf{E}[\|X_i\varepsilon_i\|\langle X_i\phi_j\rangle^2] \le \mathsf{E}[\|X_i\varepsilon_i\|^2]^{1/2}\mathsf{E}[\langle X_i\phi_j\rangle^4]^{1/2} \le C\gamma_j,$$

which implies that $\mathsf{E}[\|X_i\varepsilon_i\|\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2] = O(h_n)$. We then have

$$\mathsf{E}\left[s_{h_n}(X_0)^{-1}\left(n^{-1}\sum_{i=1}^{n}\|X_i\varepsilon_i\|\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2\right)\Big|X_0\right] = O_{\mathsf{P}}(1),$$

under Condition (A6). Therefore, we have the desired result. □

**Proposition 20.** *We suppose that* $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2} \to 0$ *as* $n \to \infty$ *and* $\sup_{j\in\mathbb{N}}\gamma_j^{-1}\|\Lambda\phi_j\|^2 < \infty$. *Then, the scaling* $s_{h_n}(X_0)$ *and* $\hat{s}_{h_n}(X_0)$ *are asymptotically equivalent in that, for any* $\eta > 0$,

$$\mathsf{P}\left(\left|\frac{\hat{s}_{h_n}(X_0)}{s_{h_n}(X_0)}-1\right| > \eta\Big|X_0\right) \xrightarrow{\mathsf{P}} 0.$$

*Proof.* We obtain the following decomposition from (3.43):

$$\begin{aligned}\left|\frac{\hat{s}_{h_n}(X_0)}{s_{h_n}(X_0)}-1\right| \leq & s_{h_n}(X_0)^{-1}|\langle(\hat{\Lambda}_{n,k_n}-\Lambda)(\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1})X_0,(\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1})X_0\rangle| \\ & + s_{h_n}(X_0)^{-1}|\langle\Lambda(\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1})X_0,(\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1})X_0\rangle| \\ & + 2s_{h_n}(X_0)^{-1}|\langle(\hat{\Lambda}_{n,k_n}-\Lambda)\Gamma_{h_n}^{-1}X_0,(\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1})X_0\rangle| \\ & + 2s_{h_n}(X_0)^{-1}|\langle\Lambda\Gamma_{h_n}^{-1}X_0,(\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1})X_0\rangle| \\ & + s_{h_n}(X_0)^{-1}|\langle(\hat{\Lambda}_{n,k_n}-\Lambda)\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\rangle|.\end{aligned} \tag{3.46}$$

We now investigate each term in the upper bound by using the previous lemmas. Before that, we observe that $\mathsf{E}[\|\tilde{\Lambda}_n-\Lambda\|_\infty^2]=O(n^{-1})$ by Theorem 2.5 of [32] since $\mathsf{E}[\|X\varepsilon\|^4]<\infty$.

By the observation that $\mathsf{E}[\|\tilde{\Lambda}_n-\Lambda\|_\infty^2]=O(n^{-1})$, the first parts of each of Lemmas 23-25, and the decomposition (3.44), we have that $\|\hat{\Lambda}_{n,k_n}-\Lambda\|_\infty \xrightarrow{\mathsf{P}} 0$ as $n\to\infty$. Then, Lemma 20 along with Lemma 17 implies that the first two terms in (3.46) converges to zero either conditionlly or unconditionally on $X_0$ if $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\to 0$.

Recall that $\mathsf{E}[\|\Gamma_{h_n}^{-1}X_0\|^2]=\sum_{j=1}^{h_n}\gamma_j^{-1}$. Since $\mathsf{E}[\|\tilde{\Lambda}_n-\Lambda\|_\infty^2]=O(n^{-1})$, we have that

$$\begin{aligned}& s_{h_n}(X_0)^{-1}|\langle(\tilde{\Lambda}_n-\Lambda)\Gamma_{h_n}^{-1}X_0,(\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1})X_0\rangle| \\ \leq & \{h_n s_{h_n}(X_0)^{-1}\}h_n^{-1}O_{\mathsf{P}}(n^{-1/2})O_{\mathsf{P}}\left(\left(\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2}\right)\|(\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1})X_0\| \\ = & \{h_n s_{h_n}(X_0)^{-1}\}O_{\mathsf{P}}\left(\left(n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2}\right)\{h_n^{-1/2}\|(\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1})X_0\|\}.\end{aligned}$$

Thus, as $n\to\infty$, if $n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}=O(1)$, which can be achieved from Condition (A5), and if $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\to 0$, Lemma 20 implies that

$$s_{h_n}(X_0)^{-1}|\langle(\tilde{\Lambda}_n-\Lambda)\Gamma_{h_n}^{-1}X_0,(\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1})X_0\rangle|$$

converges to zero either conditionlly or unconditionally on $X_0$. By the second parts of each of Lemmas 23-25, and the decomposition (3.44), the third term in (3.46) converges to zero either conditionally on $X_0$.

For the fourth term in (3.46), we first note that $\mathsf{E}[\|\Lambda\Gamma_{h_n}^{-1}X_0\|^2] = \sum_{j=1}^{h_n}\gamma_j^{-1}\|\Lambda\phi_j\|^2$ so that

$$h_n^{-1/2}\|\Lambda\Gamma_{h_n}^{-1}X_0\| = O_{\mathsf{P}}\left(\left(h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\|\Lambda\phi_j\|^2\right)^{1/2}\right).$$

Thus, as $n\to\infty$, if

$$n^{-1/2}h_n^{-1}\left(\sum_{j=1}^{h_n}\gamma_j^{-1}\|\Lambda\phi_j\|^2\right)^{1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\to 0,$$

by Lemma 20, the fourth term in (3.46) converges to zero either conditionally on $X_0$ Note that this condition is satisfied if $\sup_{j\in\mathbb{N}}\gamma_j^{-1}\|\Lambda\phi_j\|^2<\infty$ and $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\to 0$ as $n\to\infty$.

Recall that $\mathsf{E}[\|\Gamma_{h_n}^{-1}X_0\|^2] = \sum_{j=1}^{h_n}\gamma_j^{-1}$. Since $\mathsf{E}[\|\tilde{\Lambda}_n-\Lambda\|_\infty^2] = O(n^{-1})$, we have that

$$\begin{aligned}&\mathsf{E}[h_n^{-1}|\langle(\tilde{\Lambda}_n-\Lambda)\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\rangle|]\\ \leq& h_n^{-1}\mathsf{E}[\|\tilde{\Lambda}_n-\Lambda\|_\infty]\mathsf{E}[\|\Gamma_{h_n}^{-1}X_0\|^2]\\ =&O\left(n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right).\end{aligned}$$

Meanwhile, note from Jensen's inequality that

$$n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\leq n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\gamma_j^{-1}\leq\left(n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-2}\right)^{1/2}\leq\left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}\right)^{1/2}$$

Thus, under Condition (A5), we have that

$$\begin{aligned}s_{h_n}(X_0)^{-1}|\langle(\tilde{\Lambda}_n-\Lambda)\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\rangle| &= o_{\mathsf{P}}(1),\\ \mathsf{E}[s_{h_n}(X_0)^{-1}|\langle(\tilde{\Lambda}_n-\Lambda)\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\rangle||X_0] &= o_{\mathsf{P}}(1).\end{aligned}$$

Finally, by the second parts of Lemma 23, third prats of each of Lemmas 24-25, and the decomposition (3.44), the last term in (3.46) converges to zero either conditionally on $X_0$.

The above four arguments completes the proof along with the decomposition (3.46). $\square$

## 3.10 Technical details: validity of the paired bootstrap

In this section, we complete the proofs for the consistency of the paired bootstrap and the failure of a naive paired bootstrap in Section 3.3.2. The proofs require bootstrap counterparts of lemmas and the consistency of bootstrap estimator described in Sections 3.9.1-3.9.2, respectively, which are provided in Sections 3.10.1-3.10.2. The bias terms in the decomposition (3.48)-(3.49) given below are studied in Section 3.10.3 while the lemmas for the weak convergence of the variance term is proved in Section 3.10.4. In Section 3.10.5, we provide the consistency of bootstrap scaling. The propositions used to prove the failure of a naive bootstrap method are given in Section 3.10.6. Finally, Section 3.10.7 provides the proof of the failure of both modified and naive bootstrap methods.

We notice the following bias-variance decomposition of bootstrap quantity $\hat{\beta}^*_{h_n} - \hat{\beta}_{g_n}$ as

$$\hat{\beta}^*_{h_n} - \hat{\beta}_{g_n} = \hat{\beta}^*_{h_n} - \hat{\Pi}_{h_n}\hat{\beta}_{g_n} + \hat{\Pi}_{h_n}\hat{\beta}_{g_n} - \hat{\beta}_{g_n}, \tag{3.47}$$

where the non-random bias part $\hat{\Pi}_{h_n}\hat{\beta}_{g_n} - \hat{\beta}_{g_n}$ in the bootstrap world vanishes if $h_n \geq g_n$. When $h_n < g_n$, the biased quantity $\hat{\beta}^*_{h_n} - \hat{\Pi}_{h_n}\hat{\beta}_{g_n}$ can be represented with the following decomposition:

$$\begin{aligned}
\hat{\beta}^*_{h_n} - \hat{\Pi}_{h_n}\hat{\beta}_{g_n} &= (\hat{\Gamma}^*_{h_n})^{-1}\{\hat{\Gamma}^*_n\hat{\beta}_{g_n} + U^*_n - \hat{U}_{n,g_n}\} - \hat{\Pi}_{h_n}\hat{\beta}_{g_n} \\
&= (\hat{\Gamma}^*_{h_n})^{-1}\{U^*_n - \hat{U}_{n,g_n}\} + (\hat{\Pi}^*_{h_n} - \hat{\Pi}_{h_n})\hat{\beta}_{g_n} \\
&= \{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}\{U^*_n - \hat{U}_{n,g_n}\} + \Gamma^{-1}_{h_n}\{U^*_n - \hat{U}_{n,g_n}\} \\
&\quad + (\hat{\Pi}^*_{h_n} - \Pi_{h_n})(\hat{\beta}_{g_n} - \beta) + (\hat{\Pi}^*_{h_n} - \Pi_{h_n})\beta \\
&\quad - (\hat{\Pi}_{h_n} - \Pi_{h_n})(\hat{\beta}_{g_n} - \beta) - (\hat{\Pi}_{h_n} - \Pi_{h_n})\beta.
\end{aligned} \tag{3.48}$$

Here, $U^*_{n,g_n} \equiv n^{-1}\sum_{i=1}^n (X^*_i - \bar{X}^*)(\varepsilon^*_{i,g_n} - (\overline{\varepsilon^*})_{n,g_n}) = (\overline{X^*\varepsilon^*})_{n,g_n} - \bar{X}^*(\overline{\varepsilon^*})_{n,g_n}$ where $\varepsilon^*_{i,g_n} \equiv Y^*_i - \langle X^*_i, \hat{\beta}_{g_n}\rangle$ are bootstrap errors with their average $(\overline{\varepsilon^*})_{n,g_n} \equiv n^{-1}\sum_{i=1}^n \varepsilon^*_{i,g_n}$ and $(\overline{X^*\varepsilon^*})_{n,g_n} \equiv n^{-1}\sum_{i=1}^n X^*_i\varepsilon^*_{i,g_n}$, $\bar{X}^* \equiv n^{-1}\sum_{i=1}^n X^*_i$. The non-random bias $\hat{\Pi}_{h_n}\hat{\beta}_{g_n} - \hat{\beta}_{g_n}$ is expressed as

$$\begin{aligned}
&\hat{\Pi}_{h_n}\hat{\beta}_{g_n} - \hat{\beta}_{g_n} = (\hat{\Pi}_{h_n} - \Pi_{h_n} + \Pi_{h_n} - I)\hat{\beta}_{g_n} \\
=&(\hat{\Pi}_{h_n} - \Pi_{h_n})(\hat{\beta}_{g_n} - \beta) + (\hat{\Pi}_{h_n} - \Pi_{h_n})\beta + (\Pi_{h_n} - I)(\hat{\beta}_{g_n} - \beta) + (\Pi_{h_n} - I)\beta.
\end{aligned} \tag{3.49}$$

Throughout this section, we write $\mathsf{P}^* = \mathsf{P}(\cdot|\mathcal{X}_n, \mathcal{Y}_n)$ and $\mathsf{E}^*[\cdot|\mathcal{X}_n, \mathcal{Y}_n]$ for the bootstrap probability and expectation operator, respectively.

### 3.10.1 Preliminaries: perturbation theory with bootstrap quantities

In this section, we study a bootstrap counterpart of the perturbation theory similar to those in Section 3.9.1, which can be applied to our bootstrap theory. To do this, consider the following notations:

$$G_n^*(z) = (zI - \Gamma)^{-1/2}(\hat{\Gamma}_n^* - \Gamma)(zI - \Gamma)^{-1/2};$$

$$K_n^*(z) = (zI - \Gamma)^{1/2}(zI - \hat{\Gamma}_n^*)^{-1}(zI - \Gamma)^{1/2};$$

$$\mathcal{E}_j^* = (\|G_n^*(z)\|_\infty < 1/2, \forall z \in \mathcal{B}_j);$$

$$\mathcal{A}_{h_n}^* = \{\forall j \in \{1, \dots, h_n\}, |\hat{\gamma}_j^* - \gamma_j| < \delta_j/2\}.$$

**Lemma 26.** *Suppose that Condition (A2) holds. We then have that*

$$\sup_{l,k\in\mathbb{N}} \frac{\mathsf{E}[\mathsf{E}^*[\langle(\hat{\Gamma}_n^* - \Gamma)\phi_l, \phi_k\rangle^2]]}{\gamma_l\gamma_k} \leq \frac{C_1}{n} + \frac{C_2}{n^2}.$$

*Proof.* Note that

$$\langle(\hat{\Gamma}_n^* - \Gamma)\phi_l, \phi_k\rangle^2 \leq 2\langle(\tilde{\Gamma}_n^* - \Gamma)\phi_l, \phi_k\rangle^2 + 2\langle\bar{X}^*, \phi_l\rangle^2\langle\bar{X}^*, \phi_k\rangle^2, \tag{3.50}$$

where $\tilde{\Gamma}_n^* \equiv n^{-1}\sum_{i=1}^n (X_i^*)^{\otimes 2}$. We use a similar strategy to the proof of Lemma 13.

To bound the first term in (3.50), we start with the following decomposition:

$$\begin{aligned}
&\langle(\tilde{\Gamma}_n^* - \Gamma)(\phi_l), \phi_k\rangle^2 \\
=&\left\{n^{-1}\sum_{i=1}^n\langle(X_i^*)^{\otimes 2}\phi_l, \phi_k\rangle - \langle\Gamma\phi_l, \phi_k\rangle\right\}^2 = \left\{n^{-1}\sum_{i=1}^n\langle X_i^*, \phi_l\rangle\langle X_i^*, \phi_k\rangle - \langle\Gamma\phi_l, \phi_k\rangle\right\}^2 \\
=&n^{-2}\left(\sum_{i=1}^n\langle X_i^*, \phi_l\rangle\langle X_i^*, \phi_k\rangle\right)^2 - \frac{2}{n}\sum_{i=1}^n\langle X_i^*, \phi_l\rangle\langle X_i^*, \phi_k\rangle\langle\Gamma\phi_l, \phi_k\rangle + \langle\Gamma\phi_l, \phi_k\rangle^2 \\
=&n^{-2}\sum_{i=1}^n\langle X_i^*, \phi_l\rangle^2\langle X_i^*, \phi_k\rangle^2 + n^{-2}\sum_{i\neq i'}\langle X_i^*, \phi_l\rangle\langle X_i^*, \phi_k\rangle\langle X_{i'}^*, \phi_l\rangle\langle X_{i'}^*, \phi_k\rangle \\
&- \frac{2}{n}\sum_{i=1}^n\langle X_i^*, \phi_l\rangle\langle X_i^*, \phi_k\rangle\langle\Gamma\phi_l, \phi_k\rangle + \langle\Gamma\phi_l, \phi_k\rangle^2.
\end{aligned}$$

Since $X_i^*$ and $X_{i'}^*$ are independent (under $P^*$), and

$$\mathsf{E}^*[\langle X_i^*, \phi_l\rangle\langle X_i^*, \phi_k\rangle] = \mathsf{E}^*[\langle (X_1^* \otimes X_1^*)(\phi_l), \phi_k\rangle] = \langle (\mathsf{E}^*[X_1^* \otimes X_1^*])(\phi_l), \phi_k\rangle$$
$$= \langle \hat{\Gamma}_n \phi_l, \phi_k\rangle,$$

we have that

$$\begin{aligned}
&\mathsf{E}^*[\langle (\tilde{\Gamma}_n^* - \Gamma)(\phi_l), \phi_k\rangle^2] \\
=&n^{-1}\mathsf{E}^*[\langle X_1^*, \phi_l\rangle^2\langle X_1^*, \phi_k\rangle^2] + \frac{n^2-n}{n^2}\langle \hat{\Gamma}_n\phi_l, \phi_k\rangle^2 - 2\langle \hat{\Gamma}_n\phi_l, \phi_k\rangle\langle \Gamma\phi_l, \phi_k\rangle + \langle \Gamma\phi_l, \phi_k\rangle^2 \\
=&n^{-1}\left(\mathsf{E}^*[\langle X_1^*, \phi_l\rangle^2\langle X_1^*, \phi_k\rangle^2] - \langle \hat{\Gamma}_n\phi_l, \phi_k\rangle^2\right) + \langle (\hat{\Gamma}_n - \Gamma)\phi_l, \phi_k\rangle^2 \\
=&n^{-1}\left(n^{-1}\sum_{i=1}^n \langle X_i, \phi_l\rangle^2\langle X_i, \phi_k\rangle^2 - \langle \hat{\Gamma}_n\phi_l, \phi_k\rangle^2\right) + \langle (\hat{\Gamma}_n - \Gamma)\phi_l, \phi_k\rangle^2 \\
\leq&n^{-1}\left(n^{-1}\sum_{i=1}^n \langle X_i, \phi_l\rangle^2\langle X_i, \phi_k\rangle^2\right) + \langle (\hat{\Gamma}_n - \Gamma)\phi_l, \phi_k\rangle^2.
\end{aligned}$$

This implies that

$$\mathsf{E}[\mathsf{E}^*[\langle (\tilde{\Gamma}_n^* - \Gamma)(\phi_l), \phi_k\rangle^2]] \leq n^{-1}\mathsf{E}[\langle X_1, \phi_l\rangle^2\langle X_1, \phi_k\rangle^2] + \mathsf{E}[\langle (\hat{\Gamma}_n - \Gamma)\phi_l, \phi_k\rangle^2].$$

Notice that

$$\begin{aligned}
\mathsf{E}[\langle X_1, \phi_l\rangle^2\langle X_1, \phi_k\rangle^2] &\leq \gamma_l\gamma_k\mathsf{E}[\gamma_l^{-1}\langle X_1, \phi_l\rangle^2\gamma_k^{-1}\langle X_1, \phi_k\rangle^2] \\
&\leq \gamma_l\gamma_k\sqrt{\mathsf{E}[\gamma_l^{-2}\langle X_1, \phi_l\rangle^4]}\sqrt{\mathsf{E}[\gamma_l^{-2}\langle X_1, \phi_k\rangle^4]} \leq C\gamma_l\gamma_k
\end{aligned}$$

by Condition (A2), and

$$\begin{aligned}
\mathsf{E}[\langle (\hat{\Gamma}_n - \Gamma)(\phi_l), \phi_k\rangle^2] &= n^{-1}\mathsf{E}[\langle X_1, \phi_l\rangle^2\langle X_1, \phi_k\rangle^2] + \frac{n^2-n}{n^2}\langle \Gamma\phi_l, \phi_k\rangle^2 - 2\langle \Gamma\phi_l, \phi_k\rangle^2 + \langle \Gamma\phi_l, \phi_k\rangle^2 \\
&= n^{-1}\left(\mathsf{E}[\langle X_1, \phi_l\rangle^2\langle X_1, \phi_k\rangle^2] - \langle \Gamma\phi_l, \phi_k\rangle^2\right) \leq Cn^{-1}\gamma_l\gamma_k.
\end{aligned}$$

This means that $\mathsf{E}[\mathsf{E}^*[\langle (\tilde{\Gamma}_n^* - \Gamma)(\phi_l), \phi_k\rangle^2]] \leq Cn^{-1}\gamma_l\gamma_k$.

To bounde the second term in (3.50), we first consider the case of $l \neq k$. Note that

$$\langle \bar{X}^*, \phi_l\rangle^2\langle \bar{X}^*, \phi_k\rangle^2 = n^{-4}\sum_{1\leq i,i',i_0,i_0'\leq n}\langle X_i^*, \phi_l\rangle\langle X_{i_0}^*, \phi_l\rangle\langle X_{i'}^*, \phi_k\rangle\langle X_{i_0'}^*, \phi_k\rangle.$$

We investigate the values of $\mathcal{J}_{i,i',i_0,i_0'} \equiv \mathsf{E}^*[\langle X_i^*, \phi_l\rangle\langle X_{i_0}^*, \phi_l\rangle\langle X_{i'}^*, \phi_k\rangle\langle X_{i_0'}^*, \phi_k\rangle]$ depending on the quadraple $(i, i', i_0, i_0')$.

Suppose that $i, i', i_0, i_0'$ are all distinct. We then have

$$\mathcal{J}_{i,i',i_0,i_0'} = \langle \bar{X}, \phi_l\rangle^2\langle \bar{X}, \phi_k\rangle^2.$$

Then, we have

$$\begin{aligned}
&\mathsf{E}[\langle \bar{X}, \phi_l\rangle^2\langle \bar{X}, \phi_k\rangle^2] \\
=&n^{-4}\sum_{i\neq i'}\mathsf{E}[\langle X_i, \phi_l\rangle^2\langle X_{i'}, \phi_k\rangle^2] + n^{-4}\sum_{i=1}^{n}\mathsf{E}[\langle X_i, \phi_l\rangle^2\langle X_i, \phi_k\rangle^2] \\
=&n^{-4}\{(n^2-n)\gamma_l\gamma_k + n\mathsf{E}[\langle X, \phi_l\rangle^2\langle X, \phi_k\rangle^2]\} \\
\leq&n^{-4}\left\{(n^2-n)\gamma_l\gamma_k + n\gamma_l\gamma_k\sqrt{\mathsf{E}[\xi_l^4]}\sqrt{\mathsf{E}[\xi_k^4]}\right\} \\
\leq&n^{-4}\{(n^2-n)\gamma_l\gamma_k + Cn\gamma_l\gamma_k\} \\
\leq&C\frac{\gamma_l\gamma_k}{n^2}.
\end{aligned}$$

Suppose that only two of $i, i', i_0, i_0'$ are equal and the other two are distinct from the equal value. We can divide this into three cases as follows. If $i' = i_0'$ and $i \neq i_0$ are distinct from $i' = i_0'$, then

$$\mathcal{J}_{i,i',i_0,i_0'} = \langle \bar{X}, \phi_l\rangle^2\left(n^{-1}\sum_{i=1}^{n}\langle X_i, \phi_k\rangle^2\right) = \langle \bar{X}^{\otimes 2}\phi_l, \phi_l\rangle\langle \tilde{\Gamma}_n\phi_k, \phi_k\rangle.$$

Its expected value is

$$\begin{aligned}
\mathsf{E}[\mathcal{J}_{i,i',i_0,i_0'}] &= n^{-3}\mathsf{E}\left[\left(\sum_{i=1}^{n}\langle X_i, \phi_l\rangle\right)^2\left(\sum_{i'=1}^{n}\langle X_{i'}, \phi_k\rangle^2\right)\right] \\
&= n^{-3}\sum_{i=1}^{n}\mathsf{E}[\langle X_i, \phi_l\rangle^2\langle X_i, \phi_k\rangle^2] + n^{-3}\sum_{i\neq i'}\mathsf{E}[\langle X_i, \phi_l\rangle^2\langle X_{i'}, \phi_k\rangle^2] \\
&= n^{-2}\mathsf{E}[\langle X, \phi_l\rangle^2\langle X, \phi_k\rangle^2] + n^{-3}(n^2-n)\gamma_l\gamma_k \\
&\leq Cn^{-2}\gamma_l\gamma_k + n^{-3}(n^2-n)\gamma_l\gamma_k \\
&\leq C\frac{\gamma_l\gamma_k}{n}.
\end{aligned}$$

If $i = i_0$ and $i' \neq i'_0$ are distinct from $i = i_0$, then

$$\mathcal{J}_{i,i',i_0,i'_0} = \langle \bar{X}, \phi_k \rangle^2 \left( n^{-1} \sum_{i=1}^{n} \langle X_i, \phi_l \rangle^2 \right) = \langle \bar{X}^{\otimes 2} \phi_k, \phi_k \rangle \langle \tilde{\Gamma}_n \phi_l, \phi_l \rangle.$$

Similarly to the previous case, we have that $\mathsf{E}[\mathcal{J}_{i,i',i_0,i'_0}] \leq C n^{-1} \gamma_l \gamma_k$. If either $i = i'$, $i = i'_0$, $i' = i_0$, or $i_0 = i'_0$, then,

$$\begin{aligned} \mathcal{J}_{i,i',i_0,i'_0} &= \left( n^{-1} \sum_{i=1}^{n} \langle X_i, \phi_l \rangle \langle X_i, \phi_k \rangle \right) \langle \bar{X}, \phi_l \rangle \langle \bar{X}, \phi_k \rangle \\ &= \langle \tilde{\Gamma}_n \phi_l, \phi_k \rangle \langle \bar{X}^{\otimes 2} \phi_l, \phi_k \rangle. \end{aligned}$$

Its expected value is

$$\begin{aligned} \mathsf{E}[\mathcal{J}_{i,i',i_0,i'_0}] &= n^{-3} \mathsf{E} \left[ \left( \sum_{i=1}^{n} \langle X_i, \phi_l \rangle \langle X_i, \phi_k \rangle \right) \left( \sum_{i'=1}^{n} \langle X_{i'}, \phi_l \rangle \langle X_{i'}, \phi_k \rangle \right) \right] \\ &= n^{-3} \sum_{i=1}^{n} \mathsf{E}[\langle X_i, \phi_l \rangle^2 \langle X_i, \phi_k \rangle^2] + n^{-3} \sum_{i \neq i'} \mathsf{E}[\langle X_i, \phi_l \rangle \langle X_i, \phi_k \rangle \langle X_{i'}, \phi_l \rangle \langle X_{i'}, \phi_k \rangle] \\ &= n^{-2} \mathsf{E}[\langle X, \phi_l \rangle^2 \langle X, \phi_k \rangle^2] \\ &\leq C n^{-2} \gamma_l \gamma_k. \end{aligned}$$

Suppose that three of $i, i', i_0, i'_0$ are equal and the other one is distinct from the equal value. This is divided into the following two cases. If either $i = i_0 = i'$ or $i = i_0 = i'_0$, then

$$\mathcal{J}_{i,i',i_0,i'_0} = n^{-1} \sum_{i=1}^{n} \langle X_i, \phi_l \rangle^2 \langle X_i, \phi_k \rangle \langle \bar{X}, \phi_k \rangle.$$

Its expected value is

$$\begin{aligned} \mathsf{E}[\mathcal{J}_{i,i',i_0,i'_0}] &= n^{-2} \mathsf{E} \left[ \left( \sum_{i=1}^{n} \langle X_i, \phi_l \rangle^2 \langle X_i, \phi_k \rangle \right) \left( \sum_{i'=1}^{n} \langle X_{i'}, \phi_k \rangle \right) \right] \\ &= n^{-2} \sum_{i=1}^{n} \mathsf{E}[\langle X_i, \phi_l \rangle^2 \langle X_i, \phi_k \rangle^2] \\ &= n^{-1} \mathsf{E}[\langle X, \phi_l \rangle^2 \langle X, \phi_k \rangle^2] \\ &\leq C n^{-1} \gamma_l \gamma_k. \end{aligned}$$

If either $i = i' = i_0'$ or $i_0 = i' = i_0'$, then

$$\mathcal{J}_{i,i',i_0,i_0'} = n^{-1}\sum_{i=1}^{n}\langle X_i, \phi_k\rangle^2\langle X_i, \phi_l\rangle\langle \bar{X}, \phi_l\rangle.$$

In a simlar way to the previous case, we have that $\mathsf{E}[\mathcal{J}_{i,i',i_0,i_0'}] \le Cn^{-1}\gamma_l\gamma_k$.

Suppose that $i = i' = i_0 = i_0'$, then $\mathcal{J}_{i,i',i_0,i_0'} = n^{-1}\sum_{i=1}^{n}\langle X_i, \phi_k\rangle^2\langle X_i, \phi_l\rangle^2$. Hence,

$$\mathsf{E}[\mathcal{J}_{i,i',i_0,i_0'}] = \mathsf{E}[\langle X, \phi_l\rangle^2\langle X, \phi_k\rangle^2] \le C\gamma_l\gamma_k.$$

By incorporating the number of possibilities of the quadruple $(i, i', i_0, i_0')$, we have that

$$\begin{aligned}
&\mathsf{E}[\mathsf{E}^*[\langle \bar{X}^*, \phi_l\rangle^2\langle \bar{X}^*, \phi_k\rangle^2]] \\
\le& Cn^{-4}\{n\gamma_l\gamma_k + 4n^2(n^{-1}\gamma_l\gamma_k) + 2n^3(n^{-1}\gamma_l\gamma_k) + 4n^3(n^{-2}\gamma_l\gamma_k) + n^4(n^{-2}\gamma_l\gamma_k)\} \\
\le& Cn^{-2}\gamma_l\gamma_k.
\end{aligned}$$

In a similar manner, we can show the same inequality when $l = k$. □

**Lemma 27.** *Suppose that Conditions (A2)-(A3) hold. For sufficiently large $j$, we have*

$$\mathsf{E}\left[\mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j}\|G_n^*(z)\|_\infty^2\right]\right] \le Cn^{-1}(j\log j)^2.$$

*Proof.* Let $z \in \mathcal{B}_j$. Note that $z = \gamma_j + (\delta_j/2)e^{\iota\theta}$ for some $\theta \in [0, 2\pi]$ and $|z - \gamma_j| = \delta_j/2$. By bounding the sup norm by the Hilbert-Schmidt one, we have

$$\begin{aligned}
\|G_n^*(z)\|_\infty^2 \le \|G_n^*(z)\|_{HS}^2 &= \sum_{l,k\in\mathbb{N}}|\langle G_n^*(z)\phi_l, \phi_k\rangle|^2 \\
&= \sum_{l,k\in\mathbb{N}}\left|\langle(\hat{\Gamma}_n^* - \Gamma)(zI - \Gamma)^{-1/2}\phi_l, (zI - \Gamma)^{-1/2}\phi_k\rangle\right|^2 \\
&= \sum_{l,k\in\mathbb{N}}\left|\langle(\hat{\Gamma}_n^* - \Gamma)(z - \gamma_l)^{-1/2}\phi_l, (z - \gamma_k)^{-1/2}\phi_k\rangle\right|^2 \\
&= \sum_{l,k\in\mathbb{N}}\frac{\langle(\hat{\Gamma}_n^* - \Gamma)(\phi_l), \phi_k\rangle^2}{|z - \gamma_l||z - \gamma_k|}.
\end{aligned}$$

Note that for $z \in \mathcal{B}_j$ and $i \ne j$,

$$\begin{aligned}
|z - \gamma_i| = |\gamma_j - \gamma_i + (\delta_j/2)e^{\iota\theta}| &\ge |\gamma_j - \gamma_i| - \delta_j/2 \\
&\ge |\gamma_j - \gamma_i|/2 \ge \delta_j/2
\end{aligned}$$

since $\delta_j = \min\{\gamma_j - \gamma_{j+1}, \gamma_{j-1} - \gamma_j\} \le |\gamma_j - \gamma_i|$. This implies that

$$\frac{1}{|z-\gamma_l||z-\gamma_k|}\begin{cases} \le \frac{4}{|\gamma_j-\gamma_l||\gamma_j-\gamma_k|} & \text{if } l \neq j, k \neq j, \\ \le \frac{4}{\delta_j|\gamma_j-\gamma_k|} & \text{if } l = j, k \neq j, \\ = 4\delta_j^{-2} & \text{if } l = j = k. \end{cases}$$

Combining these three observations, the sum is separated into the three parts:

$$\sup_{z\in\mathcal{B}_j}\|G_n^*(z)\|_\infty^2 \le 4\left\{\sum_{l\neq j,k\neq j}\frac{\langle(\hat{\Gamma}_n^*-\Gamma)(\phi_l),\phi_k\rangle^2}{|\gamma_j-\gamma_l||\gamma_j-\gamma_k|} + \sum_{k\neq j}\frac{\langle(\hat{\Gamma}_n^*-\Gamma)(\phi_j),\phi_k\rangle^2}{\delta_j|\gamma_j-\gamma_k|} + \frac{\langle(\hat{\Gamma}_n^*-\Gamma)(\phi_j),\phi_j\rangle^2}{\delta_j^2}\right\}.$$

Then, by Lemmas 12 and 26 and, we have

$$\begin{aligned} \mathsf{E}\left[\mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j}\|G_n^*(z)\|_\infty^2\right]\right] &\le Cn^{-1}\left\{\sum_{l\neq j,k\neq j}\frac{\gamma_l\gamma_k}{|\gamma_j-\gamma_l||\gamma_j-\gamma_k|} + \sum_{k\neq j}\frac{\gamma_j\gamma_k}{\delta_j|\gamma_j-\gamma_k|} + \frac{\gamma_j^2}{\delta_j^2}\right\} \\ &= Cn^{-1}\left\{\left(\sum_{k\neq j}\frac{\gamma_k}{|\gamma_j-\gamma_k|}\right)^2 + \frac{\gamma_j}{\delta_j}\sum_{k\neq j}\frac{\gamma_k}{|\gamma_j-\gamma_k|} + \frac{\gamma_j^2}{\delta_j^2}\right\} \\ &\le Cn^{-1}\left\{(Cj\log j)^2 + (j+1)(Cj\log j) + (j+1)^2\right\} \\ &\le Cn^{-1}(j\log j)^2 \end{aligned}$$

since $\gamma_j/\delta_j = \gamma_j/(\gamma_j - \gamma_{j+1}) \le j+1$. □

**Lemma 28.** *Suppose that Conditions (A2)-(A3) hold. Then, we have for all $z \in \mathcal{B}_j$,*

$$\|K_n^*(z)\|_\infty \mathbb{I}_{\mathcal{E}_j} \le C \quad \textit{and} \quad \mathsf{E}[\mathsf{P}^*((\mathcal{E}_j^*)^c)] \le Cn^{-1/2} j\log j.$$

*Proof.* Recall the (second) resolvednt identity for $\Gamma$ and $\hat{\Gamma}_n^*$:

$$\begin{aligned} (zI-\hat{\Gamma}_n^*)^{-1} - (zI-\Gamma)^{-1} &= (zI-\hat{\Gamma}_n^*)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\Gamma)^{-1} \\ &= (zI-\Gamma)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\hat{\Gamma}_n^*)^{-1}. \end{aligned}$$

This implies that

$$
\begin{aligned}
K_n^*(z) &= (zI-\Gamma)^{1/2}(zI-\hat{\Gamma}_n^*)^{-1}(zI-\Gamma)^{1/2} \\
&= I + (zI-\Gamma)^{-1/2}(\hat{\Gamma}_n^*-\Gamma)(zI-\hat{\Gamma}_n^*)^{-1}(zI-\Gamma)^{1/2} \\
&= I + (zI-\Gamma)^{-1/2}(\hat{\Gamma}_n^*-\Gamma)(zI-\Gamma)^{-1/2}(zI-\Gamma)^{1/2}(zI-\hat{\Gamma}_n^*)^{-1}(zI-\Gamma)^{1/2} \\
&= I + G_n^*(z)K_n^*(z)
\end{aligned}
$$

and $I = K_n^*(z) - G_n^*(z)K_n^*(z) = \{I - G_n^*(z)\}K_n^*(z)$. Recall from Theorem 3.5.5 in [33] that for linear operator $T$ with $\|T\|_\infty < 1$, $I - T$ is invertible with bounded inverse and $(I-T)^{-1} = \sum_{j=0}^{\infty} T^j$. Thus,

$$
\begin{aligned}
\|K_n^*(z)\|_\infty \mathbb{I}_{\mathcal{E}_j^*} &= \left\|\{I - G_n^*(z)\}^{-1}\right\|_\infty \mathbb{I}_{\mathcal{E}_j^*} \\
&\le \sum_{j=0}^{\infty} \|G_n^*(z)\|_\infty^j \, \mathbb{I}_{\mathcal{E}_j^*} \le \sum_{j=0}^{\infty} 2^{-j} \le 2.
\end{aligned}
$$

For the second part, it follows from the Markov inequality that

$$
\begin{aligned}
\mathsf{P}^*((\mathcal{E}_j^*)^c) = \mathsf{P}^*(\|G_n^*(z)\|_\infty \ge 1/2, \exists z \in \mathcal{B}_j) &\le \mathsf{P}^*\left(\sup_{z\in\mathcal{B}_j} \|G_n^*(z)\|_\infty \ge 1/2\right) \\
&\le 2\mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j} \|G_n^*(z)\|_\infty\right] \le 2\left\{\mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j} \|G_n^*(z)\|_\infty^2\right]\right\}^{1/2}.
\end{aligned}
$$

By Lemma 27, we finally have that

$$
(\mathsf{E}[\mathsf{P}^*((\mathcal{E}_j^*)^c)])^2 \le \mathsf{E}[\mathsf{P}^*((\mathcal{E}_j^*)^c)^2] \le 2\mathsf{E}\left[\mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j} \|G_n^*(z)\|_\infty^2\right]\right] \le Cn^{-1}(j\log j)^2,
$$

i.e., $\mathsf{E}[\mathsf{P}^*((\mathcal{E}_j^*)^c)] \le Cn^{-1/2} j \log j$. □

The following lemma gives a sufficient condition (on $h_n$) under which $\hat{\mathcal{B}}_j^*$ may be replaced with $\mathcal{B}_j$ as Lemma 17 does.

**Lemma 29.**

1. *We observe that*

$$
\begin{aligned}
\hat{\Pi}_{h_n}^* - \Pi_{h_n} &= \frac{1}{2\pi\iota}\int_{\mathcal{C}_{h_n}} \{(zI-\hat{\Gamma}_n^*)^{-1} - (zI-\Gamma)^{-1}\}dz + r_{1n}^* \mathbb{I}_{(\mathcal{A}_{h_n}^*)^c}, \\
(\hat{\Gamma}_{h_n}^*)^{-1} - \Gamma_{h_n}^{-1} &= \frac{1}{2\pi\iota}\int_{\mathcal{C}_{h_n}} z^{-1}\{(zI-\hat{\Gamma}_n^*)^{-1} - (zI-\Gamma)^{-1}\}dz + r_{2n}^* \mathbb{I}_{(\mathcal{A}_{h_n}^*)^c},
\end{aligned}
$$

*where*

$$r^*_{1n} = \hat{\Pi}^*_{h_n} - \frac{1}{2\pi\iota}\int_{\mathcal{C}_{h_n}} (zI - \hat{\Gamma}^*_n)^{-1} dz,$$
$$r^*_{2n} = (\hat{\Gamma}^*_{h_n})^{-1} - \frac{1}{2\pi\iota}\int_{\mathcal{C}_{h_n}} z^{-1}(zI - \hat{\Gamma}^*_n)^{-1} dz.$$

*2. Suppose that Conditions (A2) and (A3) hold. We have that*

$$\mathsf{E}[\mathsf{P}^*((\mathcal{A}^*_{h_n})^c)] \le C_1 n^{-1/2}\sum_{j=1}^{h_n} j\log j + C_2 n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}.$$

*Proof.* This is derived in a similar manner to the proof of Lemma 17. Note that

$$\mathsf{P}^*((A^*_{h_n})^c) \le \sum_{j=1}^{k_n}\mathsf{P}^*(|\hat{\gamma}^*_j - \gamma_j| \ge \delta_j/2) \le \sum_{j=1}^{k_n}\frac{\mathsf{E}^*[|\hat{\gamma}^*_j - \gamma_j|]}{\delta_j/2}$$

and

$$|\hat{\gamma}^*_j - \gamma_j| \le |\langle(\hat{\Gamma}^*_n - \Gamma)\phi_j, \phi_j\rangle| + |(\hat{\gamma}^*_j - \gamma_j) - \langle(\hat{\Gamma}^*_n - \Gamma)\phi_j, \phi_j\rangle|.$$

We first notice by Lemma 26 that

$$\left\{\mathsf{E}[\mathsf{E}^*[|\langle(\hat{\Gamma}^*_n - \Gamma)\phi_j, \phi_j\rangle|]]\right\}^2 \le \mathsf{E}\left[\left\{\mathsf{E}^*[|\langle(\hat{\Gamma}^*_n - \Gamma)\phi_j, \phi_j\rangle|]\right\}^2\right] \le \mathsf{E}[\mathsf{E}^*[\langle(\hat{\Gamma}^*_n - \Gamma)\phi_j, \phi_j\rangle^2]]$$
$$\le Cn^{-1}\gamma_j^2,$$

i.e., $\mathsf{E}[\mathsf{E}^*[|\langle(\hat{\Gamma}^*_n - \Gamma)\phi_j, \phi_j\rangle|]] \le Cn^{-1/2}\gamma_j$.

Next, to study the approximation error, set $\hat{c}^*_j = \operatorname{sign}(\hat{\phi}^*_j, \phi_j)$. We then see that

$$\begin{aligned}\langle(\hat{\Gamma}^*_n - \Gamma)(\hat{\phi}^*_j), \hat{c}^*_j\phi_j\rangle &= \langle\hat{\Gamma}^*_n\hat{\phi}^*_j, \hat{c}^*_j\phi_j\rangle - \langle\hat{\phi}^*_j, \hat{c}^*_j\Gamma\phi_j\rangle = \langle\hat{\gamma}^*_j\hat{\phi}^*_j, \hat{c}^*_j\phi_j\rangle - \langle\hat{\phi}^*_j, \hat{c}^*_j\gamma_j\phi_j\rangle \\ &= (\hat{\gamma}^*_j - \gamma_j)\langle\hat{\phi}^*_j, \hat{c}^*_j\phi_j\rangle = (\hat{\gamma}^*_j - \gamma_j)(\langle\hat{\phi}^*_j, \hat{c}^*_j\phi_j\rangle - 1) + (\hat{\gamma}^*_j - \gamma_j) \\ &= (\hat{\gamma}^*_j - \gamma_j)(\langle\hat{\phi}^*_j, \hat{c}^*_j\phi_j - \hat{\phi}^*_j\rangle) + (\hat{\gamma}^*_j - \gamma_j),\end{aligned}$$

which implies that

$$\left|\hat{\gamma}^*_j - \gamma_j - \langle(\hat{\Gamma}^*_n - \Gamma)(\hat{\phi}^*_j), \hat{c}^*_j\phi_j\rangle\right| = |\hat{\gamma}^*_j - \gamma_j|\left|\langle\hat{\phi}^*_j, \hat{\phi}^*_j - \hat{c}^*_j\phi_j\rangle\right| \le |\hat{\gamma}^*_j - \gamma_j|\|\hat{\phi}^*_j - \hat{c}^*_j\phi_j\|.$$

We also observe that

$$\begin{aligned}
&\left|\langle(\hat{\Gamma}_n^* - \Gamma)(\phi_j), \phi_j\rangle - \langle(\hat{\Gamma}_n^* - \Gamma)(\hat{\phi}_j^*), \hat{c}_j^*\phi_j\rangle\right| \\
=&\left|\langle(\hat{\Gamma}_n^* - \Gamma)(\phi_j), \phi_j - \hat{c}_j^*\hat{\phi}_j^*\rangle + \langle(\hat{\Gamma}_n^* - \Gamma)(\phi_j), \hat{c}_j^*\hat{\phi}_j^*\rangle - \langle(\hat{\Gamma}_n^* - \Gamma)(\hat{\phi}_j^*), \hat{c}_j^*\phi_j\rangle\right| \\
=&\left|\langle(\hat{\Gamma}_n^* - \Gamma)(\phi_j), \hat{c}_j^*\phi_j - \hat{\phi}_j^*\rangle\right| \leq \|\hat{\Gamma}_n^* - \Gamma\|_\infty \|\hat{c}_j^*\phi_j - \hat{\phi}_j^*\|
\end{aligned}$$

Combining these two results, we have

$$\begin{aligned}
&\left|\hat{\gamma}_j^* - \gamma_j - \langle(\hat{\Gamma}_n^* - \Gamma)(\phi_j), \phi_j\rangle\right| \\
\leq&\left|\hat{\gamma}_j^* - \gamma_j - \langle(\hat{\Gamma}_n^* - \Gamma)(\hat{\phi}_j^*), \hat{c}_j^*\phi_j\rangle\right| + \left|\langle(\hat{\Gamma}_n^* - \Gamma)(\hat{\phi}_j^*), \hat{c}_j^*\phi_j\rangle - \langle(\hat{\Gamma}_n^* - \Gamma)(\phi_j), \phi_j\rangle\right| \\
\leq&|\hat{\gamma}_j^* - \gamma_j|\|\hat{\phi}_j^* - \hat{c}_j^*\phi_j\| + \|\hat{\Gamma}_n^* - \Gamma\|_\infty \|\hat{c}_j^*\phi_j - \hat{\phi}_j^*\| \\
\leq&C\delta_j^{-1}\|\hat{\Gamma}_n^* - \Gamma\|_\infty^2.
\end{aligned}$$

Here, we frequently used the facts that $\sup_{j\in\mathbb{N}}|\hat{\gamma}_j^* - \gamma_j| \leq \|\hat{\Gamma}_n^* - \Gamma\|_\infty$ and that $\|\hat{\phi}_j^* - \hat{c}_j^*\phi_j\| \leq C\delta_j^{-1}\|\hat{\Gamma}_n^* - \Gamma\|_\infty$, which can be obtained from Lemmas 2.2-2.3 in [32].

Meanwhile, we see that

$$\mathsf{E}^*[\|\hat{\Gamma}_n^* - \hat{\Gamma}_n\|_\infty^2] \leq 2\mathsf{E}^*[\|\tilde{\Gamma}_n^* - \hat{\Gamma}_n\|_\infty^2] + 2\mathsf{E}^*[\|\bar{X}^*\|^4] \tag{3.51}$$

where $\tilde{\Gamma}_n^* \equiv n^{-1}\sum_{i=1}^n (X_i^*)^{\otimes 2}$.

To bound the first term in (3.51), note that

$$\begin{aligned}
&\mathsf{E}^*[\|\tilde{\Gamma}_n^* - \hat{\Gamma}_n\|_\infty^2] \leq \mathsf{E}^*[\|\tilde{\Gamma}_n^* - \hat{\Gamma}_n\|_{HS}^2] \\
\leq&n^{-1}\mathsf{E}^*[\|X_1^* \otimes X_1^* - \hat{\Gamma}_n\|_{HS}^2] = n^{-1}\left(n^{-1}\sum_{i=1}^n \|X_i^{\otimes 2} - \hat{\Gamma}_n\|_{HS}^2\right) \\
\leq&n^{-2}\sum_{i=1}^n (2\|X_i^{\otimes 2} - \Gamma\|_{HS}^2 + 4\|\tilde{\Gamma}_n - \Gamma\|_{HS}^2 + 4\|\bar{X}\|^4)
\end{aligned}$$

since $X_1^* \otimes X_1^* - \hat{\Gamma}_n, \ldots, X_n^* \otimes X_n^* - \hat{\Gamma}_n$ are iid with mean zero under $P^*$. By taking expectation $\mathsf{E}$, we have that

$$\begin{aligned}
\mathsf{E}[\mathsf{E}^*[\|\tilde{\Gamma}_n^* - \hat{\Gamma}_n\|_\infty^2]] &\leq Cn^{-1}(\mathsf{E}[\|X_1^{\otimes 2} - \Gamma\|_{HS}^2] + \mathsf{E}[\|\tilde{\Gamma}_n - \Gamma\|_{HS}^2] + \mathsf{E}[\|\bar{X}\|^4]) \\
&= Cn^{-1}\{\mathrm{tr}(\mathsf{var}[X_1^{\otimes 2}]) + n^{-1}\mathsf{E}[\|X_1\|^4]\} + Cn^{-2} \leq Cn^{-1}
\end{aligned}$$

by Theorem 2.5 in [32] and the above derivation for the upper bound of $\mathsf{E}[\|\bar{X}\|^4]$. This implies that

$$\begin{aligned}\mathsf{E}[\mathsf{E}^*[\|\tilde{\Gamma}_n^* - \Gamma\|_\infty^2]] &\le 2\mathsf{E}[\mathsf{E}^*[\|\tilde{\Gamma}_n^* - \hat{\Gamma}_n\|_\infty^2]] + 4\mathsf{E}[\|\tilde{\Gamma}_n - \Gamma\|_\infty^2] + 4\mathsf{E}[\|\bar{X}\|^4]) \\ &\le Cn^{-1} + Cn^{-2} \le Cn^{-1}.\end{aligned}$$

For an upper bound of the second term in (3.51), note that

$$\begin{aligned}n^4\|\bar{X}^*\|^4 &= \left(\sum_{i=1}^n \|X_i^*\|^2 + \sum_{i\neq i'}\langle X_i, X_{i'}\rangle\right)^2 \le 2\left(\sum_{i=1}^n \|X_i^*\|^2\right)^2 + 2\left(\sum_{i\neq i'}\langle X_i, X_{i'}\rangle\right)^2 \\ &= 2\sum_{i=1}^n \|X_i^*\|^4 + 2\sum_{i\neq i'}\|X_i^*\|^2\|X_{i'}^*\|^2 + 2\sum_{i\neq i'}\langle X_i^*, X_{i'}^*\rangle^2 + 2\sum_{\substack{i\neq i', i_0\neq i_0' \\ (i,i')\neq(i_0,i_0')}}\langle X_i^*, X_{i'}^*\rangle\langle X_{i_0}^*, X_{i_0'}^*\rangle \\ &\le 2\sum_{i=1}^n \|X_i^*\|^4 + 6\sum_{i\neq i'}\|X_i^*\|^2\|X_{i'}^*\|^2 + 2\sum_{\substack{i\neq i', i_0\neq i_0' \\ (i,i')\neq(i_0,i_0') \\ (i,i')\neq(i_0',i_0)}}\langle X_i^*, X_{i'}^*\rangle\langle X_{i_0}^*, X_{i_0'}^*\rangle. \qquad (3.52)\end{aligned}$$

We first see that $\mathsf{E}^*[\|X_i^*\|^4] = n^{-1}\sum_{i=1}^n \|X_i\|^4$ and

$$\mathsf{E}^*[\|X_i^*\|^2\|X_{i'}^*\|^2] = \mathsf{E}^*[\|X_i^*\|^2]\mathsf{E}^*[\|X_{i'}^*\|^2] = \left(n^{-1}\sum_{i=1}^n \|X_i\|^2\right)^2 \le n^{-1}\sum_{i=1}^n \|X_i\|^4.$$

The third term in (3.52) should be investigated more carefully. We can divide the cases into two: the cases where just two of $(i, i', i_0, i_0')$ are equal and where all of $(i, i', i_0, i_0')$ are distinct. Suppose that just two of $(i, i', i_0, i_0')$ are equal, without loss of generality, $i = i_0$. Then, we have that

$$\begin{aligned}\mathsf{E}^*[\langle X_i^*, X_{i'}^*\rangle\langle X_{i_0}^*, X_{i_0'}^*\rangle] &= \mathsf{E}^*[\langle (X_i^*)^{\otimes 2}X_{i'}^*, X_{i_0'}^*\rangle] = \langle \tilde{\Gamma}_n\bar{X}, \bar{X}\rangle \\ &= n^{-1}\sum_{i=1}^n \langle X_i, \bar{X}\rangle^2 = n^{-3}\sum_{i=1}^n\left(\sum_{i'=1}^n \langle X_i, X_{i'}\rangle\right)^2 \\ &= n^{-3}\sum_{i=1}^n\left(\sum_{i'=1}^n \langle X_i, X_{i'}\rangle^2 + \sum_{i'\neq i_0'}\langle X_i, X_{i'}\rangle\langle X_i, X_{i_0'}\rangle\right).\end{aligned}$$

Since one of $(i, i', i_0')$ is distinct from the others and $\mathsf{E}[X] = 0$, the expected value of the second term is zero. This implies that

$$\begin{aligned}
\mathsf{E}[\mathsf{E}^*[\langle X_i^*, X_{i'}^*\rangle\langle X_{i_0}^*, X_{i_0'}^*\rangle]] &= n^{-3}\left(\sum_{i=1}^n \mathsf{E}[\|X_i\|^4] + \sum_{i\neq i'} \mathsf{E}[\langle X_i, X_{i'}\rangle^2]\right) \\
&\leq n^{-3}\{n\mathsf{E}[\|X\|^4] + (n^2 - n)\mathsf{E}[\|X\|^2]^2\} \\
&\leq Cn^{-1}.
\end{aligned}$$

Suppose that all of $(i, i', i_0, i_0')$ are distinct. Then,

$$\mathsf{E}^*[\langle X_i^*, X_{i'}^*\rangle\langle X_{i_0}^*, X_{i_0'}^*\rangle] = \|\bar{X}\|^4.$$

As above, we see that

$$\begin{aligned}
n^4\|\bar{X}\|^4 &= \left(\sum_{i=1}^n \|X_i\|^2 + \sum_{i\neq i'}\langle X_i, X_{i'}\rangle\right)^2 \leq 2\left(\sum_{i=1}^n \|X_i\|^2\right)^2 + 2\left(\sum_{i\neq i'}\langle X_i, X_{i'}\rangle\right)^2 \\
&= 2\sum_{i=1}^n \|X_i\|^4 + 2\sum_{i\neq i'}\|X_i\|^2\|X_{i'}\|^2 + 2\sum_{i\neq i'}\langle X_i, X_{i'}\rangle^2 + 2\sum_{\substack{i\neq i', i_0\neq i_0' \\ (i,i')\neq(i_0,i_0')}} \langle X_i, X_{i'}\rangle\langle X_{i_0}, X_{i_0'}\rangle \\
&\leq 2\sum_{i=1}^n \|X_i\|^4 + 6\sum_{i\neq i'}\|X_i\|^2\|X_{i'}\|^2 + 2\sum_{\substack{i\neq i', i_0\neq i_0' \\ (i,i')\neq(i_0,i_0') \\ (i,i')\neq(i_0',i_0)}} \langle X_i, X_{i'}\rangle\langle X_{i_0}, X_{i_0'}\rangle,
\end{aligned}$$

where the expected value of the third term is zero because one of $(i, i', i_0, i_0')$ is distinct from the others and $\mathsf{E}[X] = 0$. This implies that

$$\begin{aligned}
\mathsf{E}[\mathsf{E}^*[\langle X_i^*, X_{i'}^*\rangle\langle X_{i_0}^*, X_{i_0'}^*\rangle]] &\leq n^{-4}\{2n\mathsf{E}[\|X\|^4] + 6(n^2 - n)\mathsf{E}[\|X\|^2]^2\} \\
&\leq Cn^{-2}.
\end{aligned}$$

Therefore, we have that

$$\mathsf{E}[\mathsf{E}^*[\|\bar{X}^*\|^4]] \leq Cn^{-2}.$$

In summary, we see that

$$
\begin{aligned}
&\mathsf{E}[\mathsf{P}^*((A_n^*)^c)] \\
\leq &C\sum_{j=1}^{k_n}\delta_j^{-1}\mathsf{E}[\mathsf{E}^*[|\langle(\hat{\Gamma}_n^*-\Gamma)\phi_j,\phi_j\rangle|]] \\
&+C\sum_{j=1}^{k_n}\delta_j^{-1}\mathsf{E}[\mathsf{E}^*[|(\hat{\gamma}_j^*-\gamma_j)-\langle(\hat{\Gamma}_n^*-\Gamma)\phi_j,\phi_j\rangle|]] \\
\leq &Cn^{-1/2}\sum_{j=1}^{k_n}\delta_j^{-1}\gamma_j+C\mathsf{E}[\mathsf{E}^*[\|\hat{\Gamma}_n^*-\Gamma\|_\infty^2]]\sum_{j=1}^{k_n}\delta_j^{-2} \\
\leq &Cn^{-1/2}\sum_{j=1}^{k_n}j\log j+Cn^{-1}\sum_{j=1}^{k_n}\delta_j^{-2}
\end{aligned}
$$

since $\delta_j^{-1}\gamma_j \leq \gamma_j/(\gamma_j-\gamma_{j+1}) \leq j+1 \leq j\log j$. □

*Remark* 15. By Lemmas 28-29, the quantities related to $(\mathcal{E}_j^*)^c$ or $(\mathcal{A}_{h_n}^*)^c$ may be asymptotically negligible by the following arguments. See Remark S1 in the supplement of [53] for a similar discussion.

1. Let $Q_j$ be any non-negative quantity (that can be either random or fixed and can depend on $n$ or not). Note that $\mathbb{I}_{(\mathcal{E}_j^*)^c}=0$, which implies that $Q_j\mathbb{I}_{(\mathcal{E}_j^*)^c}=0$. Let $\eta>0$ be given. If $\sum_{j=1}^{h_n}Q_j\mathbb{I}_{(\mathcal{E}_j^*)^c}>\eta$, then $\sum_{j=1}^{h_n}Q_j\mathbb{I}_{(\mathcal{E}_j^*)^c}\neq 0$, and hence, there exists $j$ such that $\mathbb{I}_{(\mathcal{E}_j^*)^c}\neq 0$. We then see that

$$
\begin{aligned}
\mathsf{P}^*\left(\sum_{j=1}^{h_n}Q_j\mathbb{I}_{(\mathcal{E}_j^*)^c}>\eta\right) &\leq \sum_{j=1}^{h_n}\mathsf{P}^*(\mathbb{I}_{(\mathcal{E}_j^*)^c}\neq 0)=\sum_{j=1}^{h_n}\mathsf{P}^*((\mathcal{E}_j^*)^c) \\
&=O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n}j\log j\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathsf{P}^*\left(\sum_{j=1}^{h_n}Q_j\mathbb{I}_{(\mathcal{E}_j^*)^c}>\eta\,\Big|X_0\right) &\leq \sum_{j=1}^{h_n}\mathsf{P}^*(\mathbb{I}_{(\mathcal{E}_j^*)^c}\neq 0\,\Big|X_0)=\sum_{j=1}^{h_n}\mathsf{P}^*(\mathbb{I}_{(\mathcal{E}_j^*)^c}\neq 0)=\sum_{j=1}^{h_n}\mathsf{P}^*((\mathcal{E}_j^*)^c) \\
&=O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n}j\log j\right)
\end{aligned}
$$

by Lemma 28. Thus, any quantities multiplied by $\mathbb{I}_{(\mathcal{E}_j^*)^c}$ (or their sums) are asymptotically negligible or ignorable under the bootstrap probability $\mathsf{P}^*$ if $n^{-1/2}\sum_{j=1}^{h_n} j\log j \to 0$ as $n\to\infty$. This helps to theoretically guarantee that $\sup_{z\in\mathcal{B}_j}\|K_n^*(z)\|_\infty$ is bounded above almost surely (with upper bound not depending on $j$) based on Lemma 28.

2. Let $Q_n$ be any non-negative quantity (that can be either random or fixed and can depend on $n$ or not). Note that $\mathbb{I}_{\mathcal{A}_{h_n}^c}=0$ implies that $Q_n\mathbb{I}_{\mathcal{A}_{h_n}^c}=0$. Let $\eta>0$ be given. If $Q_n\mathbb{I}_{\mathcal{A}_{h_n}^c}>\eta$, then $Q_n\mathbb{I}_{\mathcal{A}_{h_n}^c}\neq 0$, and hence, $\mathbb{I}_{\mathcal{A}_{h_n}^c}\neq 0$. We then see that

$$\begin{aligned}\mathsf{P}^*(Q_n\mathbb{I}_{(\mathcal{A}_{h_n}^*)^c}>\eta) \le \mathsf{P}^*(Q_n\mathbb{I}_{(\mathcal{A}_{h_n}^*)^c}\neq 0) \le \mathsf{P}^*(\mathbb{I}_{(\mathcal{A}_{h_n}^*)^c}\neq 0) = \mathsf{P}^*((\mathcal{A}_{h_n}^*)^c)\\ \le O_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}+n^{-1/2}\sum_{j=1}^{h_n} j\log j\right)\end{aligned}$$

and

$$\begin{aligned}&\mathsf{P}^*(Q_n\mathbb{I}_{(\mathcal{A}_{h_n}^*)^c}>\eta\Big|X_0)\\ \le&\mathsf{P}^*(Q_n\mathbb{I}_{(\mathcal{A}_{h_n}^*)^c}\neq 0\Big|X_0)\le \mathsf{P}^*(\mathbb{I}_{(\mathcal{A}_{h_n}^*)^c}\neq 0\Big|X_0)\\ =&\mathsf{P}^*(\mathbb{I}_{(\mathcal{A}_{h_n}^*)^c}\neq 0)=\mathsf{P}^*((\mathcal{A}_{h_n}^*)^c)\\ \le& O_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}+n^{-1/2}\sum_{j=1}^{h_n} j\log j\right)\end{aligned}$$

by Lemma 29. Thus, any quantities related to $\mathbb{I}_{(\mathcal{A}_{h_n}^*)^c}$ are also asymptotically ignorable under the bootstrap probability $\mathsf{P}^*$ if $n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}\to 0$ and $n^{-1/2}\sum_{j=1}^{h_n} j\log j\to 0$ as $n\to\infty$. This aspect theoretically guarantees that the bootstrap random contour $\hat{\mathcal{C}}_{h_n}^*$ for $\hat{\Pi}_{h_n}^*$ and $(\hat{\Gamma}_{h_n}^*)^{-1}$ can be replaced with the fixed contour $\mathcal{C}_{h_n}$.

A result to deal with centering issues is provided in the following lemma.

**Lemma 30.** *Under Conditions (A1)-(A7), as $n\to\infty$, if $\|\hat{\beta}_{g_n}-\beta\|\xrightarrow{\mathsf{P}}0$, we have that* $\mathsf{E}^*[\|\bar{X}^*(\overline{\varepsilon^*})_{n,g_n}-\bar{X}(\bar{\hat{\varepsilon}})_{g_n}\|^2]=O_{\mathsf{P}}(n^{-2})$.

*Proof.* From the identity

$$\bar{X}^*(\overline{\varepsilon^*})_{n,g_n}=n^{-2}\sum_{i=1}^n X_i^*\varepsilon_{i,g_n}^*+n^{-2}\sum_{i\neq i'}X_i^*\varepsilon_{i'}^*,$$

we have that

$$\|\bar{X}^*(\overline{\varepsilon^*})_{n,g_n}\|^2 \leq 2n^{-4}\left\|\sum_{i=1}^n X_i^*\varepsilon_{i,g_n}^*\right\|^2 + 2n^{-4}\left\|\sum_{i\neq i'} X_i^*\varepsilon_{i'}^*\right\|^2$$
$$= 2n^{-4}\sum_{i=1}^n \|X_i^*\varepsilon_{i,g_n}^*\|^2 + 2n^{-4}\sum_{i\neq i'}\langle X_i^*\varepsilon_{i,g_n}^*, X_{i'}^*\varepsilon_{i'}^*\rangle \tag{3.53}$$
$$+ 2n^{-4}\sum_{i\neq i', i_0\neq i_0'}\langle X_i^*\varepsilon_{i'}^*, X_{i_0}^*\varepsilon_{i_0'}^*\rangle.$$

If $\|\hat{\beta}_{g_n} - \beta\| \xrightarrow{\mathsf{P}} 0$, taking the bootstrap expectaion $\mathsf{E}^*$ on the terms in the first two sums in (3.53) gives

$$\mathsf{E}^*[\|X_i^*\varepsilon_{i,g_n}^*\|^2] = n^{-1}\sum_{i=1}^n \|X_i\hat{\varepsilon}_{i,g_n}\|^2 = n^{-1}\sum_{i=1}^n \|X_i\varepsilon_i - X_i\langle X_i, \hat{\beta}_{g_n} - \beta\rangle\|^2$$
$$= n^{-1}\sum_{i=1}^n \|X_i\varepsilon_i\|^2 + n^{-1}\sum_{i=1}^n \|X_i^{\otimes 2}(\hat{\beta}_{g_n} - \beta)\|^2 + 2n^{-1}\sum_{i=1}^n \langle X_i\varepsilon_i, X_i^{\otimes 2}(\hat{\beta}_{g_n} - \beta)\rangle$$
$$\leq n^{-1}\sum_{i=1}^n \|X_i\varepsilon_i\|^2 + n^{-1}\sum_{i=1}^n \|X_i\|^4\|\hat{\beta}_{g_n} - \beta\|^2 + 2n^{-1}\sum_{i=1}^n \|X_i\varepsilon_i\|\|X_i\|^2\|\hat{\beta}_{g_n} - \beta\|$$
$$= \mathsf{E}[\|X\varepsilon\|^2] + o_{\mathsf{P}}(1) + \{\mathsf{E}[\|X\|^4] + o_{\mathsf{P}}(1)\}o_{\mathsf{P}}(1) + \{\mathsf{E}[\|X\varepsilon\|\|X\|^2] + o_{\mathsf{P}}(1)\}o_{\mathsf{P}}(1)$$
$$= O_{\mathsf{P}}(1), \tag{3.54}$$

where the $O_{\mathsf{P}}(1)$ term does not depend on $i$, and

$$\mathsf{E}^*[\langle X_i^*\varepsilon_{i,g_n}^*, X_{i'}^*\varepsilon_{i'}^*\rangle] = \langle \mathsf{E}^*[X_i^*\varepsilon_{i,g_n}^*], \mathsf{E}^*[X_{i'}^*\varepsilon_{i'}^*]\rangle = \|(\overline{X\hat{\varepsilon}})_{n,g_n}\|^2 = o_{\mathsf{P}}(1)$$

due to the first part of Lemma 23.

Note that

$$(\bar{\hat{\varepsilon}})_{g_n} = \bar{\varepsilon} - \langle \bar{X}, \hat{\beta}_{g_n} - \beta\rangle = O_{\mathsf{P}}(n^{-1/2}) + O_{\mathsf{P}}(n^{-1/2}\|\hat{\beta}_{g_n} - \beta\|) = O_{\mathsf{P}}(n^{-1/2}) \tag{3.55}$$

when $\|\hat{\beta}_{g_n} - \beta\| \xrightarrow{\mathsf{P}} 0$, since $\bar{X} = O_{\mathsf{P}}(n^{-1/2})$ (cf. [32], Theorem 2.3). Keeping this in mind, we now divide the cases in the third sum in (3.53) into six cases. Suppose that $i = i_0$ and $i' = i_0'$. Then,

$$\mathsf{E}^*[\langle X_i^*\varepsilon_{i'}^*, X_{i_0}^*\varepsilon_{i_0'}^*\rangle] = \mathsf{E}^*[\|X_i^*\varepsilon_{i'}^*\|^2] = \mathsf{E}^*[\|X_i^*\|^2]\mathsf{E}^*[\|\varepsilon_{i'}^*\|^2] = \left(n^{-1}\sum_{i=1}^n \|X_i\|^2\right)\left(n^{-1}\sum_{i=1}^n \hat{\varepsilon}_{i,g_n}^2\right)$$
$$= \{\mathsf{E}[\|X\|^2] + o_{\mathsf{P}}(1)\}O_{\mathsf{P}}(1) = O_{\mathsf{P}}(1)$$

since

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_{i,g_n}^2 &= n^{-1}\sum_{i=1}^{n}(\varepsilon_i - \langle X_i, \hat{\beta}_{g_n} - \beta\rangle)^2 \\
&\le n^{-1}\sum_{i=1}^{n}\varepsilon_i^2 + n^{-1}\sum_{i=1}^{n}\|X_i\|^2\|\hat{\beta}_{g_n} - \beta\|^2 + 2n^{-1}\sum_{i=1}^{n}\|X_i\varepsilon_i\|\|\hat{\beta}_{g_n} - \beta\| \\
&= \{\mathsf{E}[\varepsilon^2] + o_{\mathsf{P}}(1)\} + \{\mathsf{E}[\|X\|^2] + o_{\mathsf{P}}(1)\}o_{\mathsf{P}}(1) + \{\mathsf{E}[\|X\varepsilon\|] + o_{\mathsf{P}}(1)\}o_{\mathsf{P}}(1) \\
&= O_{\mathsf{P}}(1).
\end{aligned}
$$

Suppose that $i = i_0'$ and $i' = i_0$. Then,

$$
\mathsf{E}^*[\langle X_i^*\varepsilon_{i'}^*, X_{i_0}^*\varepsilon_{i_0'}^*\rangle] = \langle \mathsf{E}^*[X_i^*\varepsilon_{i,g_n}^*], \mathsf{E}^*[X_{i'}^*\varepsilon_{i'}^*]\rangle = \|(\overline{X\hat{\varepsilon}})_{n,g_n}\|^2 = o_{\mathsf{P}}(1)
$$

by the first part of Lemma 23. Suppose that $i = i_0$, $i'$, and $i_0'$ are distinct. Then, thanks to (3.55),

$$
\begin{aligned}
\mathsf{E}^*[\langle X_i^*\varepsilon_{i'}^*, X_{i_0}^*\varepsilon_{i_0'}^*\rangle] &= \mathsf{E}^*[\langle X_i^*\varepsilon_{i'}^*, X_i^*\varepsilon_{i_0'}^*\rangle] = \mathsf{E}^*[\|X_i^*\|^2]\mathsf{E}^*[\varepsilon_{i'}^*]\mathsf{E}^*[\varepsilon_{i_0'}^*] \\
&= \left(n^{-1}\sum_{i=1}^{n}\|X_i\|^2\right)(\bar{\hat{\varepsilon}})_{g_n}^2 = \{\mathsf{E}[\|X\|^2] + o_{\mathsf{P}}(1)\}O_{\mathsf{P}}(n^{-1}) = O_{\mathsf{P}}(n^{-1}).
\end{aligned}
$$

Suppose that $i = i_0'$, $i'$, and $i_0$ are distinct. Then, thanks to (3.55),

$$
\begin{aligned}
\mathsf{E}^*[\langle X_i^*\varepsilon_{i'}^*, X_{i_0}^*\varepsilon_{i_0'}^*\rangle] &= \mathsf{E}^*[\langle X_i^*\varepsilon_{i'}^*, X_{i_0}^*\varepsilon_{i,g_n}^*\rangle] = \langle \mathsf{E}^*[X_i^*\varepsilon_{i,g_n}^*], \mathsf{E}^*[X_{i_0}^*]\mathsf{E}^*[\varepsilon_{i'}^*]\rangle \\
&= \langle (\overline{X\hat{\varepsilon}})_{n,g_n}, \bar{X}(\bar{\hat{\varepsilon}})_{g_n}\rangle = o_{\mathsf{P}}(n^{-1})
\end{aligned}
$$

since $(\overline{X\hat{\varepsilon}})_{n,g_n} = o_{\mathsf{P}}(1)$ due to Lemma 23. Suppose that $i' = i_0$, $i$, and $i_0'$ are distinct. Then,

$$
\begin{aligned}
\mathsf{E}^*[\langle X_i^*\varepsilon_{i'}^*, X_{i_0}^*\varepsilon_{i_0'}^*\rangle] &= \mathsf{E}^*[\langle X_i^*\varepsilon_{i'}^*, X_{i'}^*\varepsilon_{i_0'}^*\rangle] = \langle \mathsf{E}^*[X_i^*]\mathsf{E}^*[\varepsilon_{i_0'}^*], \mathsf{E}^*[X_{i'}^*\varepsilon_{i'}^*]\rangle \\
&= \langle (\overline{X\hat{\varepsilon}})_{n,g_n}, \bar{X}(\bar{\hat{\varepsilon}})_{g_n}\rangle = o_{\mathsf{P}}(n^{-1})
\end{aligned}
$$

as above. Suppose that $i' = i_0'$, $i$, and $i_0$ are distinct. Then,

$$
\begin{aligned}
\mathsf{E}^*[\langle X_i^*\varepsilon_{i'}^*, X_{i_0}^*\varepsilon_{i_0'}^*\rangle] &= \mathsf{E}^*[\langle X_i^*\varepsilon_{i'}^*, X_{i_0}^*\varepsilon_{i'}^*\rangle] = \mathsf{E}^*[(\varepsilon_{i'}^*)^2]\langle \mathsf{E}^*[X_i^*], \mathsf{E}^*[X_{i_0}^*]\rangle \\
&= \left(n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_{i,g_n}^2\right)\|\bar{X}\|^2 = O_{\mathsf{P}}(1)O_{\mathsf{P}}(n^{-1}) = O_{\mathsf{P}}(n^{-1})
\end{aligned}
$$

since $n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_{i,g_n}^2 = O_{\mathsf{P}}(1)$ as above and $\bar{X} = O_{\mathsf{P}}(n^{-1/2})$. Suppose that $i, i', i_0, i_0'$ are all distinct. Then, thanks to (3.55),

$$\mathsf{E}^*[\langle X_i^*\varepsilon_{i'}^*, X_{i_0}^*\varepsilon_{i_0'}^*\rangle] = \langle \mathsf{E}^*[X_i^*]\mathsf{E}^*[\varepsilon_{i'}^*], \mathsf{E}^*[X_{i_0}^*]\mathsf{E}^*[\varepsilon_{i_0'}^*]\rangle = \|\bar{X}(\bar{\hat{\varepsilon}})_{g_n}\|^2 = O_{\mathsf{P}}(n^{-2})$$

since $\bar{X} = O_{\mathsf{P}}(n^{-1/2})$ (cf. [32], Theorem 2.3). One can summarize the above upper bounds to derive that

$$\begin{aligned}\mathsf{E}^*[\|\bar{X}^*(\overline{\varepsilon^*})_{n,g_n} - \bar{X}(\bar{\hat{\varepsilon}})_{g_n}\|^2] &\le 2\mathsf{E}^*[\|\bar{X}^*(\overline{\varepsilon^*})_{n,g_n}\|^2] + 2\|\bar{X}(\bar{\hat{\varepsilon}})_{g_n}\|^2 \\ &= O_{\mathsf{P}}(n^{-2}).\end{aligned}$$

□

### 3.10.2 Consistency of the bootstrap FPCR estimator

By using the above perturbation theory in the bootstrap world, one can derive the following bootstrap version of Lemma 20. In this subsection, we suppose that Conditions (A1)-(A5) hold.

**Lemma 31.** *As $n \to \infty$, we have the following:*

1. *If $h_n^{-1} + n^{-1/2}\sum_{j=1}^{h_n} j\log j \to 0$, then for each $\eta > 0$, we have that*
$$\mathsf{P}^*(\|(\hat{\Pi}_{h_n}^* - \Pi_{h_n})\|_\infty > \eta) \xrightarrow{\mathsf{P}} 0.$$

2. *If $h_n^{-1} + n^{-1/2}\sum_{j=1}^{h_n} \delta_j^{-1} j\log j \to 0$, then for each $\eta > 0$, we have that*
$$\mathsf{P}^*(\|(\hat{\Gamma}_{h_n}^*)^{-1} - \Gamma_{h_n}^{-1}\|_\infty > \eta) \xrightarrow{\mathsf{P}} 0.$$

3. *If $h_n^{-1} + n^{-1/2}\sum_{j=1}^{h_n} \delta_j^{1/2}(j\log j)^{3/2} \to 0$, then for each $\eta > 0$, we have that*
$$\mathsf{P}^*(\|(\hat{\Pi}_{h_n}^* - \Pi_{h_n})X_0\| > \eta | X_0) \xrightarrow{\mathsf{P}} 0.$$

4. *If $h_n^{-1} + n^{-1/2}\sum_{j=1}^{h_n} \delta_j^{-1/2}(j\log j)^{3/2} \to 0$, then for each $\eta > 0$, we have that*
$$\mathsf{P}^*(\|\{(\hat{\Gamma}_{h_n}^*)^{-1} - \Gamma_{h_n}^{-1}\}X_0\| > \eta | X_0) \xrightarrow{\mathsf{P}} 0.$$

*Proof.* Due to similarity, we prove only the third part. We observe from Lemma 14 that

$$\begin{aligned}\hat{\Pi}^*_{h_n} - \Pi_{h_n} &= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}\left\{(zI-\hat{\Gamma}^*_n)^{-1} - (zI-\Gamma)^{-1}\right\}dz + r^*_{1n}\mathbb{I}_{(\mathcal{A}^*_{h_n})^c}\\ &= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}(zI-\Gamma)^{-1/2}K^*_n(z)G^*_n(z)(zI-\Gamma)^{-1/2}dz + r^*_{1n}\mathbb{I}_{(\mathcal{A}^*_{h_n})^c}.\end{aligned}$$

This implies that $\|(\hat{\Pi}^*_{h_n} - \Pi_{h_n})X_0\| \leq C\sum_{j=1}^{h_n} A^*_j + \|r^*_{1n}\|_\infty\|X_0\|\mathbb{I}_{(\mathcal{A}^*_{h_n})^c}$, where

$$A^*_j = \int_{\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\|_\infty\|K^*_n(z)\|_\infty\|G^*_n(z)\|_\infty\|(zI-\Gamma)^{-1/2}X_0\|dz.$$

Note that for all $z\in\mathcal{B}_j$, $|z| \geq \gamma_j - \delta_j/2 \geq \gamma_j/2$. By Equation (5.3) of [33], for $z\in\mathcal{B}_j$, we have

$$\|(zI-\Gamma)^{-1/2}\|_\infty = \left(\min_{l\in\mathbb{N}}|z-\gamma_l|^{1/2}\right)^{-1} = |z-\gamma_j|^{-1/2} = (\delta_j/2)^{-1/2}.$$

Thus, by Lemma 15 and Lemma 16, we have

$$\begin{aligned}\mathsf{E}^*[A^*_j\mathbb{I}_{\mathcal{E}^*_j}|X_0] &= \int_{\mathcal{B}_j}\|zI-\Gamma\|_\infty^{-1/2}\mathsf{E}^*[\|K^*_n(z)\|_\infty\mathbb{I}_{\mathcal{E}^*_j}\|G^*_n(z)\|_\infty]\|(zI-\Gamma)^{-1/2}X_0\|dz\\ &\leq C\int_{\mathcal{B}_j}\delta_j^{-1/2}\mathsf{E}^*[\|G^*_n(z)\|_\infty]\|(zI-\Gamma)^{-1/2}X_0\|dz\\ &= C\delta_j^{1/2}\sup_{z\in\mathcal{B}_j}\mathsf{E}^*[\|G^*_n(z)\|_\infty]\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|\\ &\leq C\delta_j^{1/2}(n^{-1/2}j\log j)\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|.\end{aligned}$$

This implies that

$$\begin{aligned}\mathsf{E}[\mathsf{E}^*[A^*_j\mathbb{I}_{\mathcal{E}^*_j}|X_0]] &\leq C\delta_j^{1/2}(n^{-1/2}j\log j)\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|\right]\\ &\leq Cn^{-1/2}\delta_j^{1/2}(j\log j)^{3/2} \leq Cn^{-1/2}j\log j,\end{aligned}$$

and hence,

$$\mathsf{E}^*\left[\sum_{j=1}^{h_n}A_j\mathbb{I}_{\mathcal{E}_j}\middle|X_0\right] = O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{1/2}(j\log j)^{3/2}\right).$$

By the argument of Remark 15, under $\mathsf{P}^*$, we see that

$$\mathsf{P}^*\left(\sum_{j=1}^{h_n} A_j^* \mathbb{I}_{(\mathcal{E}_j^*)^c} > \eta \middle| X_0\right) \leq O_{\mathsf{P}}(n^{-1/2} j \log j),$$

$$\mathsf{P}^*(\|r_{2n}\|_\infty \|X_0\| \mathbb{I}_{\mathcal{A}_{h_n}^c} > \eta \big| X_0) \leq O_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}\right) + O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n} j\log j\right).$$

We thus have the desired results. ☐

**Lemma 32.** *As $n \to \infty$, if $n^{-1}\sum_{j=1}^{h_n} \delta_j^{-1/2}(j\log j)^{3/2} \to 0$, then for each $\eta > 0$, we have that*

$$\mathsf{P}^*\left(|\langle\{(\hat{\Gamma}_{h_n}^*)^{-1} - \Gamma_{h_n}^{-1}\}\{U_n^* - \hat{U}_{n,g_n}\}, X_0\rangle| > \eta \middle| X_0\right) = o_{\mathsf{P}}(1).$$

*Proof.* We observe from Lemma 29 that

$$\begin{aligned}(\hat{\Gamma}_{h_n}^*)^{-1} - \Gamma_{h_n}^{-1} &= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}\left\{(zI - \hat{\Gamma}_n^*)^{-1} - (zI-\Gamma)^{-1}\right\}dz + r_{2n}^*\mathbb{I}_{(\mathcal{A}_{h_n}^*)^c} \\ &= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}(zI-\Gamma)^{-1/2}K_n^*(z)G_n^*(z)(zI-\Gamma)^{-1/2}dz + r_{2n}^*\mathbb{I}_{(\mathcal{A}_{h_n}^*)^c}.\end{aligned}$$

This implies that

$|\langle\{(\hat{\Gamma}_{h_n}^*)^{-1} - \Gamma_{h_n}^{-1}\}\{U_n^* - (I - \hat{\Pi}_{g_n})\hat{\Delta}_n\}, X_0\rangle| \leq C\sum_{j=1}^{h_n} A_j^* + \|r_{2n}^*\|_\infty \|U_n\|\|X_0\|\mathbb{I}_{(\mathcal{A}_{h_n}^*)^c}$ where

$$A_j^* = \int_{\mathcal{B}_j}\frac{1}{|z|}\|K_n^*(z)\|_\infty\|G_n^*(z)\|_\infty\|(zI-\Gamma)^{-1/2}\|_\infty\|U_n^* - \hat{U}_{n,g_n}\|\|(zI-\Gamma)^{-1/2}X_0\|dz.$$

Thus, we have

$$\begin{aligned}&\mathsf{E}^*[A_j^*\mathbb{I}_{\mathcal{E}_j^*}|X_0] \\ =&\int_{\mathcal{B}_j}|z|^{-1}\mathsf{E}^*[\|K_n^*(z)\|_\infty\mathbb{I}_{\mathcal{E}_j^*}\|G_n^*(z)\|_\infty\|U_n^* - \hat{U}_{n,g_n}\|]\|(zI-\Gamma)^{-1/2}\|_\infty\|(zI-\Gamma)^{-1/2}X_0\|dz \\ \leq& C\delta_j^{-1/2}\int_{\mathcal{B}_j}\gamma_j^{-1}\mathsf{E}^*[\|G_n^*(z)\|_\infty\|U_n^* - \hat{U}_{n,g_n}\|]\|(zI-\Gamma)^{-1/2}X_0\|dz \\ \leq& C\delta_j^{-1/2}\int_{\mathcal{B}_j}\delta_j^{-1}(\mathsf{E}^*[\|G_n^*(z)\|_\infty^2])^{1/2}(\mathsf{E}^*[\|U_n^* - \hat{U}_{n,g_n}\|^2])^{1/2}\|(zI-\Gamma)^{-1/2}X_0\|dz \\ \leq& C\delta_j^{-1/2}\sup_{z\in\mathcal{B}_j}(\mathsf{E}^*[\|G_n^*(z)\|_\infty^2])^{1/2}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|(\mathsf{E}^*[\|U_n^* - \hat{U}_{n,g_n}\|^2])^{1/2}.\end{aligned}$$

This implies that

$$\mathsf{E}^*\left[\sum_{j=1}^{h_n} A_j^* \mathbb{I}_{\mathcal{E}_j^*}\middle| X_0\right]$$
$$\leq C(\mathsf{E}^*[\|U_n^* - \hat{U}_{n,g_n}\|^2])^{1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}\sup_{z\in\mathcal{B}_j}(\mathsf{E}^*[\|G_n^*(z)\|_\infty^2])^{1/2}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|.$$

Recall that $U_n^* \equiv n^{-1}\sum_{i=1}^n (X_i^* - \bar{X}^*)(\varepsilon_{i,g_n}^* - (\overline{\varepsilon^*})_{n,g_n}) = (\overline{X^*\varepsilon^*})_{n,g_n} - \bar{X}^*(\overline{\varepsilon^*})_{n,g_n}$ where $(\overline{X^*\varepsilon^*})_{n,g_n} \equiv n^{-1}\sum_{i=1}^n X_i^*\varepsilon_{i,g_n}^*$. Since
$\hat{U}_{n,g_n} \equiv n^{-1}\sum_{i=1}^n (X_i - \bar{X})(\hat{\varepsilon}_{i,g_n} - (\bar{\hat{\varepsilon}})_{g_n}) = (\overline{X\hat{\varepsilon}})_{n,g_n} - \bar{X}(\bar{\hat{\varepsilon}})_{g_n}$, we see that

$$\|U_n^* - \hat{U}_{n,g_n}\|^2 \leq 2\|(\overline{X^*\varepsilon^*})_{n,g_n} - (\overline{X\hat{\varepsilon}})_{n,g_n}\|^2 + 2\|\bar{X}^*(\overline{\varepsilon^*})_{n,g_n} - \bar{X}(\bar{\hat{\varepsilon}})_{g_n}\|^2. \tag{3.56}$$

As for the first term in (3.56), since $X_i^*\varepsilon_{i,g_n}^* - (\overline{X\hat{\varepsilon}})_{n,g_n}$'s are iid with mean zero under $\mathsf{P}^*$, we see that

$$\mathsf{E}^*[\|(\overline{X^*\varepsilon^*})_{n,g_n} - (\overline{X\hat{\varepsilon}})_{n,g_n}\|^2] = n^{-1}\mathsf{E}^*[\|X_i^*\varepsilon_{i,g_n}^* - (\overline{X\hat{\varepsilon}})_{n,g_n}\|^2] \leq n^{-1}\mathsf{E}^*[\|X_i^*\varepsilon_{i,g_n}^*\|^2] = O_{\mathsf{P}}(n^{-1})$$

as computed in (3.54). This means that $\mathsf{E}^*[\|U_n^* - \hat{U}_{n,g_n}\|^2] = O_{\mathsf{P}}(n^{-1})$ due to Lemma 30.

By Lemma 16 and Lemma 27, we have that

$$\mathsf{E}\left[\sum_{j=1}^{h_n}\delta_j^{-1/2}\sup_{z\in\mathcal{B}_j}(\mathsf{E}^*[\|G_n^*(z)\|_\infty^2])^{1/2}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|\right]$$
$$=\sum_{j=1}^{h_n}\delta_j^{-1/2}\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}(\mathsf{E}^*[\|G_n^*(z)\|_\infty^2])^{1/2}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|\right]$$
$$\leq\sum_{j=1}^{h_n}\delta_j^{-1/2}\left(\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\mathsf{E}^*[\|G_n^*(z)\|_\infty^2]\right]\right)^{1/2}\left(\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|^2\right]\right)^{1/2}$$
$$\leq C\sum_{j=1}^{h_n}\delta_j^{-1/2}\{n^{-1}(j\log j)^2\}^{1/2}(j\log j)^{1/2} = n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}.$$

We now have from these two bounds that

$$\mathsf{E}^*\left[\sum_{j=1}^{h_n} A_j^* \mathbb{I}_{\mathcal{E}_j^*}\middle| X_0\right] = O_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right).$$

Meanwhile, by the argument of Remark 15, we see that

$$\mathsf{P}^*\left(\sum_{j=1}^{h_n} A_j^* \mathbb{I}_{(\mathcal{E}_j^*)^c} > \eta \middle| X_0\right) = O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n} j\log j\right).$$

We thus have the desired result. □

**Lemma 33.** *As $n \to \infty$, if $\|\hat{\beta}_{g_n} - \beta\| \xrightarrow{\mathsf{P}} 0$, then we have*

$$\mathsf{E}^*[\|\Gamma_{h_n}^{-1}(U_n^* - \hat{U}_{n,g_n})\|^2] = O_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-2}\|\Lambda^{1/2}\phi_j\|^2\right) + O_{\mathsf{P}}\left(n^{-2}\sum_{j=1}^{h_n}\gamma_j^{-2}\right).$$

*Proof.* We note that

$$\Gamma_{h_n}^{-1}(U_n^* - \hat{U}_{n,g_n}) = \Gamma_{h_n}^{-1}((\overline{X^*\varepsilon^*})_{n,g_n} - \overline{X\hat{\varepsilon}}) + \Gamma_{h_n}^{-1}(\bar{X}^*(\overline{\varepsilon^*})_{n,g_n} - \bar{X}(\bar{\hat{\varepsilon}})_{g_n}).$$

The first term in the above display is bounded as follows. Note that $\|\Gamma_{h_n}^{-1}((\overline{X^*\varepsilon^*})_{n,g_n} - \overline{X\hat{\varepsilon}})\|^2 = \sum_{j=1}^{h_n}\gamma_j^{-2}\langle(\overline{X^*\varepsilon^*})_{n,g_n} - \overline{X\hat{\varepsilon}}, \phi_j\rangle^2$ and

$$\begin{aligned}\mathsf{E}^*[\langle(\overline{X^*\varepsilon^*})_{n,g_n} - \overline{X\hat{\varepsilon}}, \phi_j\rangle^2] &= n^{-1}\mathsf{E}^*[\langle X_i^*\varepsilon_{i,g_n}^* - \overline{X\hat{\varepsilon}}, \phi_j\rangle^2] = n^{-1}\mathsf{E}^*[\langle(X_i^*\varepsilon_{i,g_n}^* - \overline{X\hat{\varepsilon}})^{\otimes 2}\phi_j, \phi_j\rangle] \\ &= n^{-1}\langle\hat{\Lambda}_{n,g_n}\phi_j, \phi_j\rangle = n^{-1}\langle(\hat{\Lambda}_{n,g_n} - \Lambda)\phi_j, \phi_j\rangle + n^{-1}\langle\Lambda\phi_j, \phi_j\rangle,\end{aligned}$$

since $\{X_i^*\varepsilon_{i,g_n}^* - \overline{X\hat{\varepsilon}}\}_{i=1}^n$ are iid with mean zero under $\mathsf{P}^*$. Recall that $\|\hat{\Lambda}_{n,g_n} - \Lambda\|_\infty \xrightarrow{\mathsf{P}} 0$ if $\|\hat{\beta}_{g_n} - \beta\| \xrightarrow{\mathsf{P}} 0$ as $n \to \infty$ from the proof of Proposition 20. We therefore have that

$$\mathsf{E}^*[\|\Gamma_{h_n}^{-1}((\overline{X^*\varepsilon^*})_{n,g_n} - \overline{X\hat{\varepsilon}})\|^2] = O_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-2}\|\Lambda^{1/2}\phi_j\|^2\right).$$

Since the next term is bounded as

$$\mathsf{E}^*[\|\Gamma_{h_n}^{-1}(\bar{X}^*(\overline{\varepsilon^*})_{n,g_n} - \bar{X}(\bar{\hat{\varepsilon}})_{g_n})\|^2] \le \|\Gamma_{h_n}^{-1}\|_\infty^2\mathsf{E}^*[\|\bar{X}^*(\overline{\varepsilon^*})_{n,g_n} - \bar{X}(\bar{\hat{\varepsilon}})_{g_n}\|^2] = O_{\mathsf{P}}\left(n^{-2}\sum_{j=1}^{h_n}\gamma_j^{-2}\right)$$

due to Lemma 30, the proof is complete. □

**Theorem 10** (Consistency of the bootstrap FPCR estimator)**.** *Suppose that $h_n^{-1} + n^{-1/2}h_n^2\log h_n \to 0$ and $\|\hat{\beta}_{g_n} - \beta\| \xrightarrow{\mathsf{P}} 0$ as $n \to \infty$. Then, the bootstrap FPCR estimator $\hat{\beta}_{h_n}^*$ converges to the slope function $\beta$ in the bootstrap probability in the sense that for each $\eta > 0$,*

$$\mathsf{P}^*(\|\hat{\beta}_{h_n}^* - \beta\| > \eta) \xrightarrow{\mathsf{P}} 0.$$

*Proof.* Based on the decompositions (3.48) and (3.49), it follows from Lemmas 20, 31-33 by using the same argument in Theorem 9. □

### 3.10.3 Bias terms

Since the regressors are resampled in the paired bootstrap scheme, we have bias terms that are random in the bootstrap world more than the residual bootstrap causes, where these bootstrap bias terms are associated with $\mathcal{X}_n^* \equiv \{X_i^*\}_{i=1}^n$ as well as $\mathcal{Y}_n^* \equiv \{Y_i^*\}_{i=1}^n$, which appear in the decomposition (3.48). In what follows, we suppose that Conditions (A1)-(A6) hold. For integer $j \geq 1$, we define

$$M_{n,j} \equiv n^{-1} \sum_{l=1}^{j} \delta_l^{-1} l \log l + n^{-1/2} \left( \sum_{l=1}^{j} \gamma_l^{-2} \|\Lambda^{1/2} \phi_l\|^2 \right)^{1/2} + n^{-1/2} \sum_{l=1}^{j} l \log l \tag{3.57}$$

in general, or

$$M_{n,j} \equiv n^{-1} \sum_{l=1}^{j} \delta_l^{-1/2} (l \log l)^{3/2} + n^{-1/2} \left( \sum_{l=1}^{j} \gamma_l^{-1} \right)^{1/2} + n^{-1/2} \sum_{l=1}^{j} l \log l. \tag{3.58}$$

if $\sup_{j \in \mathbb{N}} \gamma_j^{-1} \|\Lambda^{1/2} \phi_j\|^2 < \infty$.

#### 3.10.3.1 Non-random bias terms in the bootstrap world

We treat the non-random bias terms in the bootstrap world, which are related to $\hat{\Pi}_{h_n} \hat{\beta}_{g_n} - \hat{\beta}_{g_n}$ in the decomposition (3.47). Recall that these bias terms can be non-zero only when $h_n < g_n$ since $\hat{\Pi}_{h_n} \hat{\beta}_{g_n} - \hat{\beta}_{g_n} = 0$ if $h_n \geq g_n$, so we focus on this case here.

**Proposition 21.** *As $n \to \infty$, we have*

$$\begin{aligned} &\mathsf{E}\left[ \sqrt{\frac{n}{s_{h_n}(X_0)}} |\langle (\hat{\Pi}_{h_n} - \Pi_{h_n})(\hat{\beta}_{g_n} - \beta), X_0 \rangle| \Big| X_0 \right] \\ =& O_{\mathsf{P}}\left( M_{n,g_n} h_n^{-1/2} \sum_{j=1}^{h_n} j \log j \right) + O_{\mathsf{P}}\left( n^{-1/2} h_n^{-1/2} \sqrt{\sum_{j > g_n} \beta_j^2} \sum_{j=1}^{h_n} (j \log j)^2 \right) + o_{\mathsf{P}}(1). \end{aligned}$$

*Suppose a further condition $\sup_{j \in \mathbb{N}} \gamma_j^{-1} \|\Lambda^{1/2} \phi_j\|^2 < \infty$. As $n \to \infty$, if*

$$h_n^{-1} + g_n^{-1} + n^{-1/2} h_n^{3/2} (\log h_n) g_n^2 (\log g_n) + n^{-1/2} h_n^3 (\log h_n)^2 g_n^{1/2} \to 0,$$

*then for each $\eta > 0$,*

$$\mathsf{P}\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle(\hat{\Pi}_{h_n} - \Pi_{h_n})(\hat{\beta}_{g_n} - \beta), X_0\rangle| > \eta \Big| X_0\right) \xrightarrow{\mathsf{P}} 0.$$

*Proof.* This can be proved with the same argument as Proposition S8 in the supplement of [53], and hence, the proof is omitted here. □

**Lemma 34.** *Suppose that $h_n < g_n$ with $h_n/g_n \to \tau \in (0, 1]$.*

1. *Suppose that $\tau < 1$. As $n \to \infty$, if $n^{-1/2}(g_n - h_n)^2 \to 0$,*

   *$(g_n - h_n)s_{g_n}((I - \Pi_{h_n})X_0)^{-1} = O_{\mathsf{P}}(1)$, and*

   $$\frac{(g_n - h_n)^{-1}s_{g_n}((I - \Pi_{h_n})X_0)}{h_n^{-1}s_{h_n}(X_0)} \xrightarrow{\mathsf{P}} 1,$$

   *then we have*

   $$\sup_{y\in\mathbb{R}}\left|\mathsf{P}\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\Gamma_{h_n}^{-1}U_n, X_0\rangle \le y\Big|X_0\right) - \Phi(y/\sigma^2_{\text{dim}}(\tau))\right| \xrightarrow{\mathsf{P}} 0,$$

   *where $\sigma^2_{\text{dim}}(\tau) \equiv \tau^{-1} - 1$.*

2. *We have that*

   $$\mathsf{E}\left[\left\{\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle(I - \Pi_{h_n})\Gamma_{g_n}^{-1}U_n, X_0\rangle\right\}^2\Big|X_0\right] = O_{\mathsf{P}}\left(\frac{g_n}{h_n} - 1\right) + O_{\mathsf{P}}\left(n^{-1}h_n^{-1}\sum_{j>h_n}^{g_n}\gamma_j\right).$$

   *Thus, if $\tau = 1$, we have*

   $$\mathsf{E}\left[\left\{\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle(I - \Pi_{h_n})\Gamma_{g_n}^{-1}U_n, X_0\rangle\right\}^2\Big|X_0\right] = o_{\mathsf{P}}(1).$$

*Proof.* By using the same argument as the one in Proposition 15, we can derive

$$\sup_{y\in\mathbb{R}}\left|\mathsf{P}\left(\sqrt{\frac{n}{s_{g_n}((I - \Pi_{h_n})X_0)}}\langle(I - \Pi_{h_n})\Gamma_{g_n}^{-1}U_n, X_0\rangle \le y\Big|X_0\right) - \Phi(y/\sigma^2(\tau))\right| \xrightarrow{\mathsf{P}} 0,$$

and the result follows from Slutsky theorem.

To prove the second part, note that

$$\langle(I - \Pi_{h_n})\Gamma_{g_n}^{-1}U_n, X_0\rangle^2 \le 2\langle(I - \Pi_{h_n})\Gamma_{g_n}^{-1}\overline{X\varepsilon}, X_0\rangle^2 + 2\langle(I - \Pi_{h_n})\Gamma_{g_n}^{-1}\bar{X}\bar{\varepsilon}, X_0\rangle^2.$$

The first term in the above display is bounded as follows. Since $X_i\varepsilon_i$'s are iid with mean zero, we have that

$$\begin{aligned}
&\mathsf{E}[\langle (I-\Pi_{h_n})\Gamma_{g_n}^{-1}U_n, X_0\rangle^2|X_0] \\
=&n^{-1}\mathsf{E}[\langle X_i\varepsilon_i, \Gamma_{g_n}^{-1}(I-\Pi_{h_n})X_0\rangle^2|X_0] \\
=&n^{-1}\mathsf{E}[\langle (X_i\varepsilon_i)^{\otimes 2}\Gamma_{g_n}^{-1}(I-\Pi_{h_n})X_0, \Gamma_{g_n}^{-1}(I-\Pi_{h_n})X_0\rangle|X_0] \\
=&n^{-1}\langle \Gamma\Gamma_{g_n}^{-1}(I-\Pi_{h_n})X_0, \Gamma_{g_n}^{-1}(I-\Pi_{h_n})X_0\rangle \\
=&n^{-1}\|\Gamma_{g_n}^{-1/2}(I-\Pi_{h_n})X_0\|^2 = n^{-1}\sum_{j>h_n}^{g_n}\gamma_j^{-1}\langle X_0,\phi_j\rangle^2.
\end{aligned}$$

By taking expectation again, we see that

$$\mathsf{E}[\mathsf{E}[\langle (I-\Pi_{h_n})\Gamma_{g_n}^{-1}U_n, X_0\rangle^2|X_0]] = n^{-1}\sum_{j>h_n}^{g_n}\gamma_j^{-1}\mathsf{E}[\langle X_0,\phi_j\rangle^2] = n^{-1}(g_n-h_n),$$

and hence, $\mathsf{E}[\langle (I-\Pi_{h_n})\Gamma_{g_n}^{-1}U_n, X_0\rangle^2|X_0] = O_{\mathsf{P}}(n^{-1}(g_n-h_n))$. From the assumption that $h_nV_{h_n}(X_0)^{-1} = O_{\mathsf{P}}(1)$, we have that

$$\mathsf{E}\left[\left\{\sqrt{\frac{n}{V_{h_n}(X_0)}}\langle (I-\Pi_{h_n})\Gamma_{g_n}^{-1}\overline{X\varepsilon}, X_0\rangle\right\}^2 \middle| X_0\right] = O_{\mathsf{P}}\left(\frac{g_n}{h_n}-1\right).$$

To bound the next term, by using Lemma 18, note that

$$\begin{aligned}
\mathsf{E}[\langle (I-\Pi_{h_n})\Gamma_{g_n}^{-1}\bar{X}\bar{\varepsilon}, X_0\rangle^2|X_0] &\le \mathsf{E}[\|\bar{X}\bar{\varepsilon}\|^2\|]\|\Gamma_{g_n}^{-1}(I-\Pi_{h_n})X_0\|^2 \\
&\le Cn^{-2}\|\Gamma_{g_n}^{-1}(I-\Pi_{h_n})X_0\|^2.
\end{aligned}$$

By taking expectation again, we see that

$$\mathsf{E}[\mathsf{E}[\langle (I-\Pi_{h_n})\Gamma_{g_n}^{-1}\bar{X}\bar{\varepsilon}, X_0\rangle^2|X_0]] \le Cn^{-2}\sum_{j>h_n}^{g_n}\gamma_j^{-1}.$$

From the assumption that $h_nV_{h_n}(X_0)^{-1} = O_{\mathsf{P}}(1)$, we have that

$$\mathsf{E}\left[\left\{\sqrt{\frac{n}{V_{h_n}(X_0)}}\langle (I-\Pi_{h_n})\Gamma_{g_n}^{-1}\bar{X}\bar{\varepsilon}, X_0\rangle\right\}^2 \middle| X_0\right] = O_{\mathsf{P}}\left(n^{-1}h_n^{-1}\sum_{j>h_n}^{g_n}\gamma_j^{-1}\right).$$

Thus, as $n\to\infty$, if $g_n/h_n\to 1$, we have that

$$\mathsf{E}\left[\left\{\sqrt{\frac{n}{V_{h_n}(X_0)}}\langle (I-\Pi_{h_n})\Gamma_{g_n}^{-1}U_n, X_0\rangle\right\}^2 \middle| X_0\right] = o_{\mathsf{P}}(1).$$

□

**Proposition 22.** *Suppose that $h_n < g_n$ with $h_n/g_n \to \tau \in (0,1]$. Then, we have $\sqrt{n/s_{h_n}(X_0)}\langle (I-\Pi_{h_n})(\hat\beta_{g_n}-\beta), X_0\rangle = A_n + B_n$ where the quantities $A_n$ and $B_n$ defined as*

$$A_n \equiv A_n(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle (I-\Pi_{h_n})(\hat\beta_{g_n}-\beta), X_0\rangle - B_n,$$
$$B_n \equiv B_n(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle (I-\Pi_{h_n})\Gamma_{g_n}^{-1}U_n, X_0\rangle$$

*satisfy the following.*

1.
$$\mathsf{E}[|A_n||X_0] = O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{g_n}\delta_j^{-1/2}(j\log j)^{3/2}\right) + O_{\mathsf{P}}\left(\sqrt{\frac{n}{h_n}\sum_{j>g_n}\gamma_j\beta_j^2}\right),$$
*and hence, if further, $n^{-1/2}g_n^{7/2}(\log g_n)^3 = O(1)$ and $ng_n^{-1}\sum_{j>g_n}\gamma_j\beta_j^2 = o(1)$ (cf. Proposition 19), then for each $\eta > 0$, we have $\mathsf{P}(|A_n||X_0) \xrightarrow{\mathsf{P}} 0$.*

2. *When $\tau < 1$, as $n \to \infty$, if $n^{-1/2}g_n^2 \to 0$, $(g_n - h_n)s_{g_n}((I-\Pi_{h_n})X_0)^{-1} = O_{\mathsf{P}}(1)$, and*
$$\frac{(g_n-h_n)^{-1}s_{g_n}((I-\Pi_{h_n})X_0)}{h_n^{-1}s_{h_n}(X_0)} \xrightarrow{\mathsf{P}} 1,$$
*then $\sup_{y\in\mathbb{R}}|\mathsf{P}(B_n(X_0)|X_0) - \Phi(y/\sigma^2_{\dim}(\tau))| \xrightarrow{\mathsf{P}} 0$. When $\tau = 1$, we have $\mathsf{E}[B_n(X_0)^2|X_0] \xrightarrow{\mathsf{P}} 0$ as $n \to \infty$.*

*Proof.* By using a similar story to the proof of Proposition S9 in the supplement of [53] along with Lemma 34, we have the desired result. □

#### 3.10.3.2 Random bias terms in the bootstrap world

We treat the (random) bias terms in the decomposition (3.48). We start with finding the convergence rate for the first random bias $\langle(\hat\Pi^*_{h_n} - \Pi_{h_n})\beta, X_0\rangle$ with scaling $\sqrt{n/s_{h_n}(X_0)}$. The proof goes in a similar way to the story in Section 3.9.3.1.

By applying the second resolvent identity (Lemma 14) twice, we have

$$
\begin{aligned}
&(zI-\hat{\Gamma}_n^*)^{-1}-(zI-\Gamma)^{-1}\\
=&\{(zI-\Gamma)^{-1}+(zI-\hat{\Gamma}_n^*)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\Gamma)^{-1}\}(\hat{\Gamma}_n^*-\Gamma)(zI-\Gamma)^{-1}\\
=&(zI-\Gamma)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\Gamma)^{-1}+(zI-\hat{\Gamma}_n^*)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\Gamma)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\Gamma)^{-1}\\
=&(zI-\Gamma)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\Gamma)^{-1}+(zI-\Gamma)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\Gamma)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\hat{\Gamma}_n^*)^{-1}
\end{aligned}
$$

since all quantities are symmetric. This implies that

$$
\begin{aligned}
\hat{\Pi}_{h_n}^*-\Pi_{h_n}&=\frac{1}{2\pi\iota}\int_{\mathcal{C}_{h_n}}\left\{(zI-\hat{\Gamma}_n^*)^{-1}-(zI-\Gamma)^{-1}\right\}dz+r_{1n}\mathbb{I}_{(\mathcal{A}_{h_n}^*)^c}\\
&=\mathcal{S}_n^*+\mathcal{R}_n^*+r_{1n}^*\mathbb{I}_{(\mathcal{A}_{h_n}^*)^c}
\end{aligned}
\tag{3.59}
$$

where

$$
\begin{aligned}
\mathcal{S}_n^*&=\frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}(zI-\Gamma)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\Gamma)^{-1}dz,\\
\mathcal{R}_n^*&=\frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}(zI-\Gamma)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\Gamma)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\hat{\Gamma}_n^*)^{-1}dz,
\end{aligned}
$$

and $r_{1n}^*=\hat{\Pi}_{h_n}^*-\frac{1}{2\pi\iota}\int_{\mathcal{C}_{h_n}}(zI-\hat{\Gamma}_n^*)^{-1}dz$ (cf. Lemma 29). We will see the convergences of $\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\mathcal{S}_n^*\beta,X_0\rangle$ and $\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\mathcal{R}_n^*\beta,X_0\rangle$ (based on the decomposition (3.59)) to zero, respectively, under the bootstrap probability $\mathsf{P}^*$.

**Lemma 35.** *As $n\to\infty$, if $h_n\to\infty$, we have $\frac{n}{h_n}\mathsf{E}[\mathsf{E}^*[\langle\mathcal{S}_n^*\beta,X_0\rangle^2|X_0]]=o(1)$, which implies that*

$$
\mathsf{E}^*\left[\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\mathcal{S}_n^*\beta,X_0\rangle\right)^2\Big|X_0\right]=o_{\mathsf{P}}(1).
$$

*Proof.* Notice from the expansion of $X_0$ that

$$
\begin{aligned}
\mathsf{E}^*[\langle\mathcal{S}_n^*\beta,X_0\rangle^2|X_0]&=\mathsf{E}^*\left[\left(\sum_{l=1}^{\infty}\langle X_0,\phi_l\rangle\langle\mathcal{S}_n^*\beta,\phi_l\rangle\right)^2\Big|X_0\right]\\
=&\sum_{l=1}^{\infty}\mathsf{E}^*[\langle X_0,\phi_l\rangle^2\langle\mathcal{S}_n^*\beta,\phi_l\rangle^2|X_0]+\sum_{l\neq l'}\mathsf{E}^*[\langle X_0,\phi_l\rangle\langle\mathcal{S}_n^*\beta,\phi_{l'}\rangle\langle X_0,\phi_{l'}\rangle\langle\mathcal{S}_n^*\beta,\phi_{l'}\rangle|X_0]\\
=&\sum_{l=1}^{\infty}\langle X_0,\phi_l\rangle^2\mathsf{E}^*[\langle\mathcal{S}_n^*\beta,\phi_l\rangle^2|X_0]+\sum_{l\neq l'}\langle X_0,\phi_l\rangle\langle X_0,\phi_{l'}\rangle\mathsf{E}^*[\langle\mathcal{S}_n^*\beta,\phi_{l'}\rangle\langle\mathcal{S}_n^*\beta,\phi_{l'}\rangle|X_0].
\end{aligned}
$$

Since $X_0$ is independent of $\mathcal{X}_n = \{X_i\}_{i=1}^n$ and $\{\langle X_0, \phi_l\rangle : l \in \mathbb{N}\}$ is an uncorrelated sequence with mean zero, this implies that

$$\begin{aligned}
&\mathsf{E}[\mathsf{E}^*[\langle \mathcal{S}_n^*\beta, X_0\rangle^2 | X_0]] \\
=&\sum_{l=1}^{\infty} \mathsf{E}[\langle X_0, \phi_l\rangle^2]\mathsf{E}[\mathsf{E}^*[\langle \mathcal{S}_n^*\beta, \phi_l\rangle^2 | X_0]] \\
&+\sum_{l\neq l'} \mathsf{E}[\langle X_0, \phi_l\rangle\langle X_0, \phi_{l'}\rangle]\mathsf{E}[\mathsf{E}^*[\langle \mathcal{S}_n^*\beta, \phi_{l'}\rangle\langle \mathcal{S}_n^*\beta, \phi_{l'}\rangle | X_0]] \\
=&\sum_{l=1}^{\infty} \gamma_l \mathsf{E}[\mathsf{E}^*[\langle \mathcal{S}_n^*\beta, \phi_l\rangle^2]].
\end{aligned}$$

Write $\beta_j = \langle \beta, \phi_j\rangle$ for the projection of the slope function $\beta$ onto $\phi_j$ for each $j \in \mathbb{N}$. From the basis expansion of $\beta = \sum_{l'=1}^{\infty} \beta_{l'}\phi_{l'}$, for each $l \in \mathbb{N}$, we see that

$$\langle \mathcal{S}_n^*\beta, \phi_l\rangle^2 = \left(\sum_{l'=1}^{\infty} \beta_{l'}\langle \mathcal{S}_n^*\phi_l, \phi_{l'}\rangle\right)^2.$$

To explicitly compute $\langle \mathcal{S}_n^*\phi_l, \phi_{l'}\rangle$, note that

$$\begin{aligned}
\langle \mathcal{S}_n^*\phi_l, \phi_{l'}\rangle &= \sum_{j=1}^{h_n} \frac{1}{2\pi\iota}\int_{\mathcal{B}_j} \langle (zI-\Gamma)^{-1}(\hat{\Gamma}_n^* - \Gamma)(zI-\Gamma)^{-1}\phi_l, \phi_{l'}\rangle dz \\
&= \langle (\hat{\Gamma}_n^* - \Gamma)\phi_l, \phi_{l'}\rangle \sum_{j=1}^{h_n} \frac{1}{2\pi\iota}\int_{\mathcal{B}_j} \frac{1}{z-\gamma_l}\frac{1}{z-\gamma_{l'}} dz
\end{aligned}$$

since $(zI-\Gamma)^{-1} = \sum_{l=1}^{\infty}(z-\gamma_l)^{-1}\pi_l$ (cf. [33], Equation (5.2)). By using the contour integral theory in complex analysis (cf. [1], Chapter 4), one can show that

$$\langle \mathcal{S}_n^*\phi_l, \phi_{l'}\rangle = \begin{cases} \frac{\langle (\hat{\Gamma}_n^* - \Gamma)\phi_l, \phi_{l'}\rangle}{\gamma_l - \gamma_{l'}} & \text{if } l \le h_n < l', \\ \frac{\langle (\hat{\Gamma}_n^* - \Gamma)\phi_l, \phi_{l'}\rangle}{\gamma_{l'} - \gamma_l} & \text{if } l' \le h_n < l, \\ 0 & \text{otherwise} \end{cases}$$

as the proof of Proposition 17.

We now investigate $\mathsf{E}[\mathsf{E}^*[\langle \mathcal{S}_n^*\beta, X_0\rangle^2]]$ depending on $l$. Suppose that $l \le h_n$ first. We then have

$$
\begin{aligned}
\langle \mathcal{S}_n^*\beta, \phi_l\rangle^2 &= \left(\sum_{l'>h_n}^{\infty} \beta_{l'} \frac{\langle (\hat{\Gamma}_n^* - \Gamma)\phi_l, \phi_{l'}\rangle}{\gamma_l - \gamma_{l'}}\right)^2 \\
\le & 2\left(\sum_{l'>h_n}^{\infty} \beta_{l'} \frac{\langle (\hat{\Gamma}_n^* - \tilde{\Gamma}_n)\phi_l, \phi_{l'}\rangle}{\gamma_l - \gamma_{l'}}\right)^2 + 2\left(\sum_{l'>h_n}^{\infty} \beta_{l'} \frac{\langle (\tilde{\Gamma}_n - \Gamma)\phi_l, \phi_{l'}\rangle}{\gamma_l - \gamma_{l'}}\right)^2 \\
\le & 4\left(\sum_{l'>h_n}^{\infty} \beta_{l'} \frac{\langle (\tilde{\Gamma}_n^* - \tilde{\Gamma}_n)\phi_l, \phi_{l'}\rangle}{\gamma_l - \gamma_{l'}}\right)^2 + 4\left(\sum_{l'>h_n}^{\infty} \beta_{l'} \frac{\langle \bar{X}^*, \phi_l\rangle\langle \bar{X}^*, \phi_{l'}\rangle}{\gamma_l - \gamma_{l'}}\right)^2 \\
& + 2\left(\sum_{l'>h_n}^{\infty} \beta_{l'} \frac{\langle (\tilde{\Gamma}_n - \Gamma)\phi_l, \phi_{l'}\rangle}{\gamma_l - \gamma_{l'}}\right)^2
\end{aligned}
\tag{3.60}
$$

where $\tilde{\Gamma}_n \equiv n^{-1}\sum_{i=1}^n X_i^{\otimes 2}$ and $\tilde{\Gamma}_n^* \equiv n^{-1}\sum_{i=1}^n (X_i^*)^{\otimes 2}$.

By taking the bootstrap expectation $\mathsf{E}^*$ for the first term in (3.60), we have

$$
\begin{aligned}
&\mathsf{E}^*\left[\left(\sum_{l'>h_n}^{\infty} \beta_{l'} \frac{\langle (\tilde{\Gamma}_n^* - \tilde{\Gamma}_n)\phi_l, \phi_{l'}\rangle}{\gamma_l - \gamma_{l'}}\right)^2 \Big| X_0\right] \\
=&\mathsf{E}^*\left[\left(n^{-1}\sum_{i=1}^n\sum_{l'>h_n}^{\infty} \beta_{l'} \frac{\langle (X_i^*\otimes X_i^* - \tilde{\Gamma}_n)\phi_l, \phi_{l'}\rangle}{\gamma_l - \gamma_{l'}}\right)^2 \Big| X_0\right] \\
=&n^{-1}\mathsf{E}^*\left[\left(\sum_{l'>h_n}^{\infty} \beta_{l'} \frac{\langle (X_1^*\otimes X_1^* - \tilde{\Gamma}_n)\phi_l, \phi_{l'}\rangle}{\gamma_l - \gamma_{l'}}\right)^2 \Big| X_0\right] \\
=&n^{-1}\left\{n^{-1}\sum_{i=1}^n\left(\sum_{l'>h_n}^{\infty} \beta_{l'} \frac{\langle (X_i^{\otimes 2} - \tilde{\Gamma}_n)\phi_l, \phi_{l'}\rangle}{\gamma_l - \gamma_{l'}}\right)^2\right\} \\
\le&2n^{-2}\sum_{i=1}^n\left(\sum_{l'>h_n}^{\infty} \beta_{l'} \frac{\langle (X_i^{\otimes 2} - \Gamma)\phi_l, \phi_{l'}\rangle}{\gamma_l - \gamma_{l'}}\right)^2 + 2n^{-1}\left(\sum_{l'>h_n}^{\infty} \beta_{l'} \frac{\langle (\tilde{\Gamma}_n - \Gamma)\phi_l, \phi_{l'}\rangle}{\gamma_l - \gamma_{l'}}\right)^2
\end{aligned}
$$

since $X_i^*\otimes X_i^* - \tilde{\Gamma}_n$'s are iid with mean zero under $P^*$. By taking the (original) expectation $\mathsf{E}$, we now have that

$$
\begin{aligned}
&\mathsf{E}\left[\mathsf{E}^*\left[\left(\sum_{l'>h_n}^{\infty} \beta_{l'} \frac{\langle (\tilde{\Gamma}_n^* - \tilde{\Gamma}_n)\phi_l, \phi_{l'}\rangle}{\gamma_l - \gamma_{l'}}\right)^2 \Big| X_0\right]\right] \\
\le&2n^{-1}\mathsf{E}\left[\left(\sum_{l'>h_n}^{\infty} \beta_{l'} \frac{\langle (X^{\otimes 2} - \Gamma)\phi_l, \phi_{l'}\rangle}{\gamma_l - \gamma_{l'}}\right)^2\right] + 2n^{-1}\mathsf{E}\left[\left(\sum_{l'>h_n}^{\infty} \beta_{l'} \frac{\langle (\tilde{\Gamma}_n - \Gamma)\phi_l, \phi_{l'}\rangle}{\gamma_l - \gamma_{l'}}\right)^2\right].
\end{aligned}
$$

As above, the third term in (3.60) is bounded as

$$\begin{aligned}\mathsf{E}\left[\left(\sum_{l'>h_n}^{\infty}\beta_{l'}\frac{\langle(\tilde{\Gamma}_n-\Gamma)\phi_l,\phi_{l'}\rangle}{\gamma_l-\gamma_{l'}}\right)^2\right] =&\mathsf{E}\left[\left(n^{-1}\sum_{i=1}^{n}\sum_{l'>h_n}^{\infty}\beta_{l'}\frac{\langle(X_i^{\otimes 2}-\Gamma)\phi_l,\phi_{l'}\rangle}{\gamma_l-\gamma_{l'}}\right)^2\right]\\ =&n^{-1}\mathsf{E}\left[\left(\sum_{l'>h_n}^{\infty}\beta_{l'}\frac{\langle(X^{\otimes 2}-\Gamma)\phi_l,\phi_{l'}\rangle}{\gamma_l-\gamma_{l'}}\right)^2\right]\end{aligned}$$

since $X_i^{\otimes 2}-\Gamma$'s are iid with mean zero. The upper bound for $\mathsf{E}\left[\left(\sum_{l'>h_n}^{\infty}\beta_{l'}\frac{\langle(X^{\otimes 2}-\Gamma)\phi_l,\phi_{l'}\rangle}{\gamma_l-\gamma_{l'}}\right)^2\right]$ is given as

$$\begin{aligned}\mathsf{E}\left[\left(\sum_{l'>h_n}\beta_{l'}\frac{\langle X,\phi_l\rangle\langle X,\phi_{l'}\rangle}{\gamma_l-\gamma_{l'}}\right)^2\right] &= \sum_{l',l_0'>h_n}\beta_{l'}\beta_{l_0'}\frac{\mathsf{E}[\langle X,\phi_l\rangle^2\langle X,\phi_{l'}\rangle\langle X,\phi_{l_0'}\rangle]}{(\gamma_l-\gamma_{l'})(\gamma_l-\gamma_{l_0'})}\\ &\le C\sum_{l',l_0'>h_n}\beta_{l'}\beta_{l_0'}\frac{\gamma_l\gamma_{l'}^{1/2}\gamma_{l_0'}^{1/2}}{(\gamma_l-\gamma_{l'})(\gamma_l-\gamma_{l_0'})}\\ &\le C\left(\sum_{l'>h_n}|\beta_{l'}|\frac{\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_l-\gamma_{l'}}\right)^2.\end{aligned}$$

The second term in (3.60) is bounded in a more complicated way. We first note that

$$\begin{aligned}&\left(\sum_{l'>h_n}\beta_{l'}\frac{\langle\bar{X}^*,\phi_l\rangle\langle\bar{X}^*,\phi_{l'}\rangle}{\gamma_l-\gamma_{l'}}\right)^2\\ \le&2n^{-4}\left(\sum_{i=1}^{n}\sum_{l'>h_n}\frac{\beta_{l'}}{\gamma_l-\gamma_{l'}}\langle X_i^*,\phi_l\rangle\langle X_i^*,\phi_{l'}\rangle\right)^2\\ &+2n^{-4}\left(\sum_{i\neq i'}\sum_{l'>h_n}\frac{\beta_{l'}}{\gamma_l-\gamma_{l'}}\langle X_i^*,\phi_l\rangle\langle X_{i'}^*,\phi_{l'}\rangle\right)^2.\end{aligned} \tag{3.61}$$

After taking the bootstrap expectation $\mathsf{E}^*$, the first term in (3.61) can be expanded as follows:

$$\begin{aligned}&\mathsf{E}^*\left[\left(\sum_{i=1}^{n}\sum_{l'>h_n}\frac{\beta_{l'}}{\gamma_l-\gamma_{l'}}\langle X_i^*,\phi_l\rangle\langle X_i^*,\phi_{l'}\rangle\right)^2\right]\\ =&\sum_{i=1}^{n}\sum_{l',l_0'>h_n}\frac{\beta_{l'}}{\gamma_l-\gamma_{l'}}\frac{\beta_{l_0'}}{\gamma_l-\gamma_{l_0'}}\mathsf{E}^*[\langle X_i^*,\phi_l\rangle^2\langle X_i^*,\phi_{l'}\rangle\langle X_i^*,\phi_{l_0'}\rangle]\\ &+\sum_{i\neq i_0}\sum_{l',l_0'>h_n}\frac{\beta_{l'}}{\gamma_l-\gamma_{l'}}\frac{\beta_{l_0'}}{\gamma_l-\gamma_{l_0'}}\mathsf{E}^*[\langle X_i^*,\phi_l\rangle\langle X_{i_0}^*,\phi_l\rangle\langle X_i^*,\phi_{l'}\rangle\langle X_{i_0}^*,\phi_{l_0'}\rangle].\end{aligned}$$

Here, we note that

$$
\begin{aligned}
&\mathsf{E}[\mathsf{E}^*[\langle X_i^*, \phi_l\rangle\langle X_{i_0}^*, \phi_l\rangle\langle X_i^*, \phi_{l'}\rangle\langle X_{i_0}^*, \phi_{l_0'}\rangle]] \\
=&\mathsf{E}[\mathsf{E}^*[\langle X_i^*, \phi_l\rangle\langle X_i^*, \phi_{l'}\rangle|X_0]\mathsf{E}^*[\langle X_{i_0}^*, \phi_l\rangle\langle X_{i_0}^*, \phi_{l_0'}\rangle]] \\
=&\mathsf{E}\left[\left(n^{-1}\sum_{i=1}^n\langle X_i, \phi_l\rangle\langle X_i, \phi_{l'}\rangle\right)\left(n^{-1}\sum_{i=1}^n\langle X_i, \phi_l\rangle\langle X_i, \phi_{l_0'}\rangle\right)\right] \\
=&n^{-2}\sum_{i=1}^n\mathsf{E}[\langle X_i, \phi_l\rangle\langle X_i, \phi_{l'}\rangle\langle X_i, \phi_l\rangle\langle X_i, \phi_{l_0'}\rangle] \\
&+n^{-2}\sum_{i\neq i'}\mathsf{E}[\langle X_i, \phi_l\rangle\langle X_i, \phi_{l'}\rangle\langle X_{i'}, \phi_l\rangle\langle X_{i'}, \phi_{l_0'}\rangle] \\
=&n^{-1}\mathsf{E}[\langle X, \phi_l\rangle^2\langle X, \phi_{l'}\rangle\langle X, \phi_{l_0'}\rangle] \\
&+n^{-2}\sum_{i\neq i'}\mathsf{E}[\langle X_i, \phi_l\rangle\langle X_i, \phi_{l'}\rangle]\mathsf{E}[\langle X_{i'}, \phi_l\rangle\langle X_{i'}, \phi_{l_0'}\rangle] \\
=&n^{-1}\mathsf{E}[\langle X, \phi_l\rangle^2\langle X, \phi_{l'}\rangle\langle X, \phi_{l_0'}\rangle]
\end{aligned}
$$

since $X_1, \ldots, X_n$ are independent, $l \leq h_n < l', l_0'$, and the FPC scores are uncorrelated with mean zero, which implies that $\mathsf{E}[\langle X_i, \phi_l\rangle\langle X_i, \phi_{l'}\rangle]\mathsf{E}[\langle X_{i'}, \phi_l\rangle\langle X_{i'}, \phi_{l_0'}\rangle] = 0$. In addition, we have

$$
\begin{aligned}
&\mathsf{E}[\mathsf{E}^*[\langle X_i^*, \phi_l\rangle^2\langle X_i^*, \phi_{l'}\rangle\langle X_i^*, \phi_{l_0'}\rangle]] \\
=&\mathsf{E}\left[n^{-1}\sum_{i=1}^n\langle X_i, \phi_l\rangle^2\langle X_i, \phi_{l'}\rangle\langle X_i, \phi_{l_0'}\rangle\right] \\
=&\mathsf{E}[\langle X, \phi_l\rangle^2\langle X, \phi_{l'}\rangle\langle X, \phi_{l_0'}\rangle].
\end{aligned}
$$

Note that $\mathsf{E}[\langle X, \phi_l\rangle^2\langle X, \phi_{l'}\rangle\langle X, \phi_{l_0'}\rangle] \leq C\gamma_l\gamma_{l'}^{1/2}\gamma_{l_0'}^{1/2}$ since

$$
\begin{aligned}
&\sup_{l,l',l_0'\in\mathbb{N}}\gamma_l^{-1}\gamma_{l'}^{-1/2}\gamma_{l_0'}^{-1/2}\mathsf{E}[\langle X, \phi_l\rangle^2\langle X, \phi_{l'}\rangle\langle X, \phi_{l_0'}\rangle] \\
\leq&\sup_{l\in\mathbb{N}}\mathsf{E}[\gamma_l^{-2}\langle X, \phi_l\rangle^4]^{1/2}\sup_{l'\in\mathbb{N}}\mathsf{E}[\gamma_{l'}^{-2}\langle X, \phi_{l'}\rangle^4]^{1/4}\sup_{l_0'\in\mathbb{N}}\mathsf{E}[\gamma_{l_0'}^{-2}\langle X, \phi_{l_0'}\rangle^4]^{1/4} < \infty
\end{aligned}
$$

by Cauchy-Schwarz inequality and Condition (A2). This implies that

$$
\mathsf{E}\left[\mathsf{E}^*\left[\left(\sum_{i=1}^n\sum_{l'>h_n}\frac{\beta_{l'}}{\gamma_l-\gamma_{l'}}\langle X_i^*, \phi_l\rangle\langle X_i^*, \phi_{l'}\rangle\right)^2\right]\right] \leq Cn\left(\sum_{l'>h_n}\beta_{l'}\frac{\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_l-\gamma_{l'}}\right)^2.
$$

To bound the second term in (3.61), note that

$$\mathsf{E}^*\left[\left(\sum_{i\neq i'}\sum_{l'>h_n}\frac{\beta_{l'}}{\gamma_l-\gamma_{l'}}\langle X_i^*,\phi_l\rangle\langle X_{i'}^*,\phi_{l'}\rangle\right)^2\right]$$
$$=\sum_{i\neq i',i_0\neq i_0'}\sum_{l',l_0'>h_n}\frac{\beta_{l'}}{\gamma_l-\gamma_{l'}}\frac{\beta_{l_0'}}{\gamma_l-\gamma_{l_0'}}\mathsf{E}^*[\langle X_i^*,\phi_l\rangle\langle X_{i_0}^*,\phi_l\rangle\langle X_{i'}^*,\phi_{l'}\rangle\langle X_{i_0'}^*,\phi_{l_0'}\rangle].$$

This sum is divided into the following four cases. Suppose that $(i,i')=(i_0,i_0')$, where the number of cases is $n^2-n$. Then,

$$\mathsf{E}[\mathsf{E}^*[\langle X_i^*,\phi_l\rangle\langle X_{i_0}^*,\phi_l\rangle\langle X_{i'}^*,\phi_{l'}\rangle\langle X_{i_0'}^*,\phi_{l_0'}\rangle]]$$
$$=\mathsf{E}[\mathsf{E}^*[\langle X_i^*,\phi_l\rangle^2\langle X_{i'}^*,\phi_{l'}\rangle\langle X_{i'}^*,\phi_{l_0'}\rangle]]$$
$$=\mathsf{E}[\mathsf{E}^*[\langle X_i^*,\phi_l\rangle^2]\mathsf{E}^*[\langle X_{i'}^*,\phi_{l'}\rangle\langle X_{i'}^*,\phi_{l_0'}\rangle]]$$
$$=\mathsf{E}\left[\left(n^{-1}\sum_{i=1}^n\langle X_i,\phi_l\rangle^2\right)\left(n^{-1}\sum_{i=1}^n\langle X_i,\phi_{l'}\rangle\langle X_i,\phi_{l_0'}\rangle\right)\right]$$
$$=n^{-2}\sum_{i=1}^n\mathsf{E}[\langle X_i,\phi_l\rangle^2\langle X_i,\phi_{l'}\rangle\langle X_i,\phi_{l_0'}\rangle]+n^{-2}\sum_{i\neq i'}\mathsf{E}[\langle X_i,\phi_l\rangle^2\langle X_{i'},\phi_{l'}\rangle\langle X_{i'},\phi_{l_0'}]$$
$$=n^{-1}\mathsf{E}[\langle X,\phi_l\rangle^2\langle X,\phi_{l'}\rangle\langle X,\phi_{l_0'}\rangle]+n^{-2}\sum_{i\neq i'}\mathsf{E}[\langle X_i,\phi_l\rangle^2]\mathsf{E}[\langle X_{i'},\phi_{l'}\rangle\langle X_{i'},\phi_{l_0'}]$$
$$=n^{-1}\mathsf{E}[\langle X,\phi_l\rangle^2\langle X,\phi_{l'}\rangle\langle X,\phi_{l_0'}\rangle]+(1-n^{-1})\gamma_l\gamma_{l'}\mathbb{I}(l'=l_0').$$

Secondly, suppose that $i=i_0'$ and $i'=i_0$, where the number of cases is $n^2-n$. Then,

$$\mathsf{E}[\mathsf{E}^*[\langle X_i^*,\phi_l\rangle\langle X_{i_0}^*,\phi_l\rangle\langle X_{i'}^*,\phi_{l'}\rangle\langle X_{i_0'}^*,\phi_{l_0'}\rangle]]$$
$$=\mathsf{E}[\mathsf{E}^*[\langle X_i^*,\phi_l\rangle\langle X_i^*,\phi_{l_0'}\rangle]\mathsf{E}[\langle X_{i'}^*,\phi_l\rangle\langle X_{i'}^*,\phi_{l'}\rangle]]$$
$$=\mathsf{E}\left[\left(n^{-1}\sum_{i=1}^n\langle X_i,\phi_l\rangle\langle X_i,\phi_{l_0'}\right)\left(n^{-1}\sum_{i=1}^n\langle X_i,\phi_l\rangle\langle X_i,\phi_{l'}\rangle\right)\right]$$
$$=n^{-2}\sum_{i=1}^n\mathsf{E}[\langle X_i,\phi_l\rangle^2\langle X_i,\phi_{l_0'}\rangle\langle X_i,\phi_{l'}\rangle]+n^{-2}\sum_{i\neq i'}\mathsf{E}[\langle X_i,\phi_l\rangle\langle X_i,\phi_{l_0'}\rangle\langle X_{i'},\phi_l\rangle\langle X_{i'},\phi_{l'}\rangle]$$
$$=n^{-1}\mathsf{E}[\langle X,\phi_l\rangle^2\langle X,\phi_{l_0'}\rangle\langle X,\phi_{l'}\rangle]+n^{-2}\sum_{i\neq i'}\mathsf{E}[\langle X_i,\phi_l\rangle\langle X_i,\phi_{l_0'}\rangle]\mathsf{E}[\langle X_{i'},\phi_l\rangle\langle X_{i'},\phi_{l'}\rangle]$$
$$=n^{-1}\mathsf{E}[\langle X,\phi_l\rangle^2\langle X,\phi_{l_0'}\rangle\langle X,\phi_{l'}\rangle]$$

since $X_1, \ldots, X_n$ are independent, $l \le h_n < l', l_0'$, and the FPC scores are uncorrelated with mean zero, which implies that $\mathsf{E}[\langle X_i, \phi_l\rangle\langle X_i, \phi_{l_0'}\rangle]\mathsf{E}[\langle X_{i'}, \phi_l\rangle\langle X_{i'}, \phi_{l'}\rangle] = 0$. Next, we suppose that three of $(i, i', i_0, i_0')$ are distinct with $i \neq i'$ and $i_0 \neq i_0'$, where the number of cases is $4n(n-1)(n-2)$. Then, we see that

$$\begin{aligned}&\mathsf{E}^*[\langle X_i^*, \phi_l\rangle\langle X_{i_0}^*, \phi_l\rangle\langle X_{i'}^*, \phi_{l'}\rangle\langle X_{i_0'}^*, \phi_{l_0'}\rangle]\\ =&\begin{cases}\left(n^{-1}\sum_{i=1}^n\langle X_i, \phi_l\rangle^2\right)\left(n^{-1}\sum_{i=1}^n\langle X_i, \phi_{l'}\rangle\right)\left(n^{-1}\sum_{i=1}^n\langle X_i, \phi_{l_0'}\rangle\right) & \text{if } i = i_0\\ \left(n^{-1}\sum_{i=1}^n\langle X_i, \phi_l\rangle\langle X_i, \phi_{l_0'}\rangle\right)\left(n^{-1}\sum_{i=1}^n\langle X_i, \phi_l\rangle\right)\left(n^{-1}\sum_{i=1}^n\langle X_i, \phi_{l'}\rangle\right) & \text{if } i = i_0'\\ \left(n^{-1}\sum_{i=1}^n\langle X_i, \phi_l\rangle\right)\left(n^{-1}\sum_{i=1}^n\langle X_i, \phi_l\rangle\langle X_i, \phi_{l'}\rangle\right)\left(n^{-1}\sum_{i=1}^n\langle X_i, \phi_{l_0'}\rangle\right) & \text{if } i' = i_0\\ \left(n^{-1}\sum_{i=1}^n\langle X_i, \phi_l\rangle\right)^2\left(n^{-1}\sum_{i=1}^n\langle X_i, \phi_{l'}\rangle\langle X_i, \phi_{l_0'}\rangle\right) & \text{if } i' = i_0'.\end{cases}\end{aligned}$$

One can show that

$$\mathsf{E}[\mathsf{E}^*[\langle X_i^*, \phi_l\rangle\langle X_{i_0}^*, \phi_l\rangle\langle X_{i'}^*, \phi_{l'}\rangle\langle X_{i_0'}^*, \phi_{l_0'}\rangle]] = n^{-2}\mathsf{E}[\langle X, \phi_l\rangle^2\langle X, \phi_{l_0'}\rangle\langle X, \phi_{l'}\rangle]$$

if either $i' = i_0$ or $i = i_0'$ and

$$\begin{aligned}&\mathsf{E}[\mathsf{E}^*[\langle X_i^*, \phi_l\rangle\langle X_{i_0}^*, \phi_l\rangle\langle X_{i'}^*, \phi_{l'}\rangle\langle X_{i_0'}^*, \phi_{l_0'}\rangle]]\\ =&n^{-2}\mathsf{E}[\langle X, \phi_l\rangle^2\langle X, \phi_{l_0'}\rangle\langle X, \phi_{l'}\rangle] + n^{-1}(1-n^{-1})\gamma_l\gamma_{l'}\mathbb{I}(l' = l_0')\end{aligned}$$

if either $i = i_0$ or $i' = i_0'$. As the last case, suppose that all of $(i, i', i_0, i_0')$ are distinct, where the number of cases is $n(n-1)(n-2)(n-3)$. Then, we see that

$$\begin{aligned}&\mathsf{E}^*[\langle X_i^*, \phi_l\rangle\langle X_{i_0}^*, \phi_l\rangle\langle X_{i'}^*, \phi_{l'}\rangle\langle X_{i_0'}^*, \phi_{l_0'}\rangle]\\ =&\left(n^{-1}\sum_{i=1}^n\langle X_i, \phi_l\rangle\right)^2\left(n^{-1}\sum_{i=1}^n\langle X_i, \phi_{l'}\rangle\right)\left(n^{-1}\sum_{i=1}^n\langle X_i, \phi_{l_0'}\rangle\right),\end{aligned}$$

and hence,

$$\begin{aligned}&\mathsf{E}[\mathsf{E}^*[\langle X_i^*, \phi_l\rangle\langle X_{i_0}^*, \phi_l\rangle\langle X_{i'}^*, \phi_{l'}\rangle\langle X_{i_0'}^*, \phi_{l_0'}\rangle]]\\ =&n^{-3}\mathsf{E}[\langle X, \phi_l\rangle^2\langle X, \phi_{l'}\rangle\langle X, \phi_{l_0'}\rangle] + n^{-2}(1-n^{-1})\gamma_l\gamma_{l'}\mathbb{I}(l' = l_0').\end{aligned}$$

Note that the total number of cases of $(i, i', i_0, i_0')$ with $i \neq i'$ and $i_0 \neq i_0'$ is

$$2(n^2 - n) + 4n(n-1)(n-2) + n(n-1)(n-2)(n-3) = (n^2 - n)^2.$$

By summarizing these upper bounds, we have

$$\begin{aligned}
&\mathsf{E}\left[\mathsf{E}^*\left[\left(\sum_{i\neq i'}\sum_{l'>h_n}\frac{\beta_{l'}}{\gamma_l-\gamma_{l'}}\langle X_i^*,\phi_l\rangle\langle X_{i'}^*,\phi_{l'}\rangle\right)^2\right]\right]\\
=&\sum_{l',l_0'>h_n}\frac{\beta_{l'}}{\gamma_l-\gamma_{l'}}\frac{\beta_{l_0'}}{\gamma_l-\gamma_{l_0'}}\left\{\frac{2(n^2-n)}{n}+\frac{4n(n-1)(n-2)}{n^2}+\frac{n(n-1)(n-2)(n-3)}{n^3}\right\}\\
&\qquad\times\mathsf{E}[\langle X,\phi_l\rangle^2\langle X,\phi_{l'}\rangle\langle X,\phi_{l_0'}\rangle]\\
&+\sum_{l'>h_n}\left(\frac{\beta_{l'}}{\gamma_l-\gamma_{l'}}\right)^2\left\{(n^2-n)+\frac{2n(n-1)(n-2)}{n}+\frac{n(n-1)(n-2)(n-3)}{n^2}\right\}(1-n^{-1})\gamma_l\gamma_{l'}\\
\leq& Cn\sum_{l',l_0'>h_n}\frac{|\beta_{l'}|}{\gamma_l-\gamma_{l'}}\frac{|\beta_{l_0'}|}{\gamma_l-\gamma_{l_0'}}\gamma_l\gamma_{l'}^{1/2}\gamma_{l_0'}^{1/2}+Cn^2\sum_{l'>h_n}\left(\frac{\beta_{l'}\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_l-\gamma_{l'}}\right)^2\\
\leq& Cn^2\left(\sum_{l'>h_n}\frac{|\beta_{l'}|\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_l-\gamma_{l'}}\right)^2.
\end{aligned}$$

Thus, the third term in (3.61) is bounded above as

$$\mathsf{E}\left[\mathsf{E}^*\left[\left(\sum_{l'>h_n}\beta_{l'}\frac{\langle\bar{X}^*,\phi_l\rangle\langle\bar{X}^*,\phi_{l'}\rangle}{\gamma_l-\gamma_{l'}}\right)^2\middle|X_0\right]\right]\leq Cn^{-2}\left(\sum_{l'>h_n}|\beta_{l'}|\frac{\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_l-\gamma_{l'}}\right)^2.$$

In summary, if $l \leq h_n$, we obtain

$$\mathsf{E}[\mathsf{E}^*[\langle\mathcal{S}_n^*\beta,\phi_l\rangle^2|X_0]]\leq Cn^{-1}\left(\sum_{l'>h_n}|\beta_{l'}|\frac{\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_l-\gamma_{l'}}\right)^2.$$

A similar compuation can apply to the case of $l > h_n$ so that

$$\mathsf{E}[\mathsf{E}^*[\langle\mathcal{S}_n^*\beta,X_0\rangle^2|X_0]]\leq Cn^{-1}\sum_{l=1}^{h_n}\gamma_l\left(\sum_{l'>h_n}|\beta_{l'}|\frac{\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_l-\gamma_{l'}}\right)^2+Cn^{-1}\sum_{l>h_n}\gamma_l\left(\sum_{l'=1}^{h_n}|\beta_{l'}|\frac{\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_l-\gamma_{l'}}\right)^2. \tag{3.62}$$

Following the same truncation technique in the proof of Proposition 2 in [12], one can derive that

$$\mathsf{E}\left[\mathsf{E}^*\left[\frac{n}{h_n}\langle\mathcal{S}_n^*\beta,X_0\rangle^2\middle|X_0\right]\right]=o(1). \tag{3.63}$$

This implies that

$$\mathsf{E}^*\left[\frac{n}{s_{h_n}(X_0)}\langle \mathcal{S}_n^*\beta, X_0\rangle^2\middle|X_0\right] = \{h_n s_{h_n}(X_0)^{-1}\}\mathsf{E}^*\left[\frac{n}{h_n}\langle \mathcal{S}_n^*\beta, X_0\rangle^2\middle|X_0\right] = O_{\mathsf{P}}(1)o_{\mathsf{P}}(1) = o_{\mathsf{P}}(1),$$

which conclude that

$$\mathsf{P}^*\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle \mathcal{S}_n^*\beta, X_0\rangle| > \eta\middle|X_0\right) = o_{\mathsf{P}}(1).$$

□

**Lemma 36.** *As $n\to\infty$, if $h_n^{-1} + n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}(j\log j)^2 \to 0$, we have that*

$$\mathsf{P}^*\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle \mathcal{R}_n^*\beta, X_0\rangle| > \eta\middle|X_0\right) = o_{\mathsf{P}}(1).$$

*Proof.* We observe that

$$\begin{aligned}
\mathcal{R}_n^* &= \sum_{j=1}^{h_n}\frac{1}{2\pi\iota}\int_{\mathcal{B}_j}(zI-\Gamma)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\Gamma)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\hat{\Gamma}_n^*)^{-1}dz\\
&= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}(zI-\Gamma)^{-1/2}G_n^*(z)(zI-\Gamma)^{-1/2}(\hat{\Gamma}_n^*-\Gamma)(zI-\hat{\Gamma}_n^*)^{-1}dz\\
&= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}(zI-\Gamma)^{-1/2}G_n^*(z)^2(zI-\Gamma)^{1/2}(zI-\hat{\Gamma}_n^*)^{-1}dz\\
&= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}(zI-\Gamma)^{-1/2}G_n^*(z)^2K_n^*(z)(zI-\Gamma)^{-1/2}dz.
\end{aligned}$$

This implies that $|\langle \mathcal{R}_n^*\beta, X_0\rangle| \le C\sum_{j=1}^{h_n}A_j^*$ where

$$A_j^* = \int_{\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|\|G_n^*(z)\|_\infty^2\|K_n^*(z)\|_\infty\|(zI-\Gamma)^{-1/2}\beta\|dz.$$

Thus, by Lemma 28, we have

$$\begin{aligned}
\mathsf{E}^*[A_j^*\mathbb{I}_{\mathcal{E}_j^*}|X_0] &= \int_{\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|\mathsf{E}^*\left[\|G_n^*(z)\|_\infty^2\|K_n^*(z)\|_\infty\mathbb{I}_{\mathcal{E}_j^*}\right]\|(zI-\Gamma)^{-1/2}\beta\|dz.\\
&\le C\int_{\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|\mathsf{E}^*[\|G_n^*(z)\|_\infty^2]\|(zI-\Gamma)^{-1/2}\beta\|dz.
\end{aligned}$$

Note that for all $z\in\mathcal{B}_j$, $|z|\ge\gamma_j-\delta_j/2\ge\gamma_j/2$. By Equation (5.3) of [33], for $z\in\mathcal{B}_j$, we have

$$\|(zI-\Gamma)^{-1/2}\|_\infty = \left(\min_{l\in\mathbb{N}}|z-\gamma_l|^{1/2}\right)^{-1} = |z-\gamma_j|^{-1/2} = (\delta_j/2)^{-1/2}.$$

By Lemma 16 and Lemma 27, we see that

$$\mathsf{E}[\mathsf{E}^*[A_j^*\mathbb{I}_{\mathcal{E}_j^*}|X_0]] \le C\delta_j \mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|\right]\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\mathsf{E}^*[\|G_n^*(z)\|_\infty^2\right]\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\beta\|$$
$$\le C\delta_j(j\log j)^{1/2}\{n^{-1}(j\log j)^2\}\delta_j^{-1/2} = Cn^{-1}\delta_j^{1/2}(j\log j)^{5/2} \le Cn^{-1}(j\log j)^2$$

since $\delta_j \le \gamma_j \le C(j\log j)^{-1}$. We therefore conclude that
$\mathsf{E}\left[\mathsf{E}^*\left[\sum_{j=1}^{h_n} A_j^*\mathbb{I}_{\mathcal{E}_j^*}\Big|X_0\right]\right] \le Cn^{-1}\sum_{j=1}^{h_n}(j\log j)^2$, which implies that

$$\mathsf{E}^*\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}\sum_{j=1}^{h_n}A_j^*\mathbb{I}_{\mathcal{E}_j^*}\Big|X_0\right] = O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}(j\log j)^2\right).$$

Meanwhile, by the argument of Remark 15, we see that

$$\mathsf{P}^*\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}\sum_{j=1}^{h_n}A_j^*\mathbb{I}_{(\mathcal{E}_j^*)^c} > \eta\Big|X_0\right) = O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n}j\log j\right).$$

We thus have the desired result. □

A bootstrap version of Proposition 17 is given as follows.

**Proposition 23.** *As $n\to\infty$, if $h_n^{-1} + n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}(j\log j)^2 \to 0$, then for each $\eta > 0$, we have that*

$$\mathsf{P}^*\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle(\hat{\Pi}_{h_n}^* - \Pi_{h_n})\beta, X_0\rangle| > \eta\Big|X_0\right) = o_{\mathsf{P}}(1).$$

*Proof.* By the argument of Remark 15, we see that

$$\mathsf{P}^*\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}\|\beta\|\|r_{1n}^*\|_\infty\|X_0\|\mathbb{I}_{(\mathcal{A}_{h_n}^*)^c} > \eta\Big|X_0\right) \le C_1n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2} + C_2n^{-1/2}\sum_{j=1}^{h_n}j\log j.$$

Thus, under Condition (A5), we have that

$$\mathsf{P}^*\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle r_{1n}^*\mathbb{I}_{(\mathcal{A}_{h_n}^*)^c}\beta, X_0\rangle| > \eta\Big|X_0\right) \xrightarrow{\mathsf{P}} 0,$$

and by Lemmas 35-36 and the decomposition (3.59), we have that

$$\mathsf{P}^*\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle(\hat{\Pi}_{h_n}^* - \Pi_{h_n})\beta, X_0\rangle| > \eta\Big|X_0\right) = o_{\mathsf{P}}(1).$$

□

We now state and prove a bootstrap version of Proposition 18.

**Proposition 24.** *As $n \to \infty$, if $h_n^{-1} + n^{-1/2} h_n^{-1/2} \sum_{j=1}^{h_n} \delta_j^{-1/2} (j \log j)^{3/2} \to 0$, then for each $\eta > 0$, we have that*

$$\mathsf{P}^* \left( \sqrt{\frac{n}{s_{h_n}(X_0)}} |\langle \{ (\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n} \} \{ U_n^* - \hat{U}_{n,g_n} \}, X_0 \rangle | > \eta \Big| X_0 \right) = o_{\mathsf{P}}(1).$$

*Proof.* We observe from Lemma 29 that

$$\begin{aligned}(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n} &= \frac{1}{2\pi\iota} \sum_{j=1}^{h_n} \int_{\mathcal{B}_j} z^{-1} \left\{ (zI - \hat{\Gamma}^*_n)^{-1} - (zI - \Gamma)^{-1} \right\} dz + r^*_{2n} \mathbb{I}_{(\mathcal{A}^*_{h_n})^c} \\ &= \frac{1}{2\pi\iota} \sum_{j=1}^{h_n} \int_{\mathcal{B}_j} z^{-1} (zI - \Gamma)^{-1/2} K_n^*(z) G_n^*(z) (zI - \Gamma)^{-1/2} dz + r^*_{2n} \mathbb{I}_{(\mathcal{A}^*_{h_n})^c}.\end{aligned}$$

This implies that
$|\langle \{ (\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n} \} (U_n^* - \hat{U}_{n,g_n}), X_0 \rangle| \leq C \sum_{j=1}^{h_n} A_j^* + \|r^*_{2n}\|_\infty \|U_n^* - \hat{U}_{n,g_n}\| \|X_0\| \mathbb{I}_{(\mathcal{A}^*_{h_n})^c}$ where

$$A_j^* = \int_{\mathcal{B}_j} \frac{1}{|z|} \|K_n^*(z)\|_\infty \|G_n^*(z)\|_\infty \|(zI - \Gamma)^{-1/2}\|_\infty \|U_n^* - \hat{U}_{n,g_n}\| \|(zI - \Gamma)^{-1/2} X_0\| dz.$$

Thus, we have

$$\begin{aligned}&\mathsf{E}^*[A_j^* \mathbb{I}_{\mathcal{E}_j^*} | X_0] \\ =&\int_{\mathcal{B}_j} |z|^{-1} \mathsf{E}^*[\|K_n^*(z)\|_\infty \mathbb{I}_{\mathcal{E}_j^*} \|G_n^*(z)\|_\infty \|U_n^* - \hat{U}_{n,g_n}\|] \|(zI - \Gamma)^{-1/2}\|_\infty \|(zI - \Gamma)^{-1/2} X_0\| dz \\ \leq& C \delta_j^{-1/2} \int_{\mathcal{B}_j} \gamma_j^{-1} \mathsf{E}^*[\|G_n^*(z)\|_\infty \|U_n^* - \hat{U}_{n,g_n}\|] \|(zI - \Gamma)^{-1/2} X_0\| dz \\ \leq& C \delta_j^{-1/2} \int_{\mathcal{B}_j} \delta_j^{-1} (\mathsf{E}^*[\|G_n^*(z)\|^2_\infty])^{1/2} (\mathsf{E}^*[\|U_n^* - \hat{U}_{n,g_n}\|^2])^{1/2} \|(zI - \Gamma)^{-1/2} X_0\| dz \\ \leq& C \delta_j^{-1/2} \sup_{z \in \mathcal{B}_j} (\mathsf{E}^*[\|G_n^*(z)\|^2_\infty])^{1/2} \sup_{z \in \mathcal{B}_j} \|(zI - \Gamma)^{-1/2} X_0\| (\mathsf{E}^*[\|U_n^* - \hat{U}_{n,g_n}\|^2])^{1/2}.\end{aligned}$$

This implies that

$$\begin{aligned}&\mathsf{E}^* \left[ \sum_{j=1}^{h_n} A_j^* \mathbb{I}_{\mathcal{E}_j^*} \Big| X_0 \right] \\ \leq& C (\mathsf{E}^*[\|U_n^* - \hat{U}_{n,g_n}\|^2])^{1/2} \sum_{j=1}^{h_n} \delta_j^{-1/2} \sup_{z \in \mathcal{B}_j} (\mathsf{E}^*[\|G_n^*(z)\|^2_\infty])^{1/2} \sup_{z \in \mathcal{B}_j} \|(zI - \Gamma)^{-1/2} X_0\|.\end{aligned}$$

To bound the term $\mathsf{E}^*[\|U_n^* - \hat{U}_{n,g_n}\|^2]$ in the preceding display, recall that $U_n^* \equiv n^{-1}\sum_{i=1}^n (X_i^* - \bar{X}^*)(\varepsilon_{i,g_n}^* - (\overline{\varepsilon^*})_{n,g_n}) = (\overline{X^*\varepsilon^*})_{n,g_n} - \bar{X}^*(\overline{\varepsilon^*})_{n,g_n}$ where $(\overline{X^*\varepsilon^*})_{n,g_n} \equiv n^{-1}\sum_{i=1}^n X_i^*\varepsilon_{i,g_n}^*$. Since $\hat{U}_{n,g_n} \equiv n^{-1}\sum_{i=1}^n (X_i - \bar{X})(\hat{\varepsilon}_{i,g_n} - (\bar{\hat{\varepsilon}})_{g_n}) = (\overline{X\hat{\varepsilon}})_{n,g_n} - \bar{X}(\bar{\hat{\varepsilon}})_{g_n}$, we see that

$$\|U_n^* - \hat{U}_{n,g_n}\|^2 \leq 2\|(\overline{X^*\varepsilon^*})_{n,g_n} - (\overline{X\hat{\varepsilon}})_{n,g_n}\|^2 + 2\|\bar{X}^*(\overline{\varepsilon^*})_{n,g_n} - \bar{X}(\bar{\hat{\varepsilon}})_{g_n}\|^2. \tag{3.64}$$

As for the first term in (3.64), since $X_i^*\varepsilon_{i,g_n}^* - (\overline{X\hat{\varepsilon}})_{n,g_n}$'s are iid with mean zero under $\mathsf{P}^*$, we see that

$$\begin{aligned}\mathsf{E}^*[\|(\overline{X^*\varepsilon^*})_{n,g_n} - (\overline{X\hat{\varepsilon}})_{n,g_n}\|^2] &= n^{-1}\mathsf{E}^*[\|X_i^*\varepsilon_{i,g_n}^* - (\overline{X\hat{\varepsilon}})_{n,g_n}\|^2] \\ &\leq 2n^{-1}(\mathsf{E}^*[\|X_i^*\varepsilon_{i,g_n}^*\|^2] + \|(\overline{X\hat{\varepsilon}})_{n,g_n}\|^2) = O_{\mathsf{P}}(n^{-1})\end{aligned}$$

as computed in (3.54) and Lemma 23. This means that $\mathsf{E}^*[\|U_n^* - \hat{U}_{n,g_n}\|^2] = O_{\mathsf{P}}(n^{-1})$ due to Lemma 30.

Next, by Lemma 16 and Lemma 27, we have that

$$\begin{aligned}&\mathsf{E}\left[\sum_{j=1}^{h_n}\delta_j^{-1/2}\sup_{z\in\mathcal{B}_j}(\mathsf{E}^*[\|G_n^*(z)\|_\infty^2])^{1/2}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|\right] \\ =&\sum_{j=1}^{h_n}\delta_j^{-1/2}\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}(\mathsf{E}^*[\|G_n^*(z)\|_\infty^2])^{1/2}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|\right] \\ \leq&\sum_{j=1}^{h_n}\delta_j^{-1/2}\left(\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\mathsf{E}^*[\|G_n^*(z)\|_\infty^2]\right]\right)^{1/2}\left(\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|^2\right]\right)^{1/2} \\ \leq& C\sum_{j=1}^{h_n}\delta_j^{-1/2}\{n^{-1}(j\log j)^2\}^{1/2}(j\log j)^{1/2} = n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}.\end{aligned}$$

We now have from these two bounds that

$$\mathsf{E}^*\left[\sum_{j=1}^{h_n}A_j^*\mathbb{I}_{\mathcal{E}_j^*}\middle|X_0\right] = O_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right),$$

and hence,

$$\mathsf{E}^*\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}\sum_{j=1}^{h_n} A_j^* \mathbb{I}_{\mathcal{E}_j^*}\right] = \sqrt{h_n s_{h_n}(X_0)^{-1}}\sqrt{\frac{n}{h_n}}\mathsf{E}^*\left[\sum_{j=1}^{h_n} A_j^* \mathbb{I}_{\mathcal{E}_j^*}\right]$$
$$= O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right).$$

Meanwhile, by the argument of Remark 15, we see that

$$\mathsf{P}^*\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}\sum_{j=1}^{h_n} A_j^* \mathbb{I}_{(\mathcal{E}_j^*)^c} > \eta \Big| X_0\right) = O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n} j\log j\right).$$

We thus have the desired result. □

The following proposition is a bootstrap version of Proposition 21.

**Proposition 25.** *As $n \to \infty$, we have*

$$\mathsf{E}^*\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle(\hat{\Pi}_{h_n}^* - \Pi_{h_n})(\hat{\beta}_{g_n} - \beta), X_0\rangle|\Big|X_0\right]$$
$$= O_{\mathsf{P}}\left(M_{n,g_n}h_n^{-1/2}\sum_{j=1}^{h_n} j\log j\right) + O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sqrt{\sum_{j>g_n}\beta_j^2\sum_{j=1}^{h_n}(j\log j)^2}\right) + o_{\mathsf{P}}(1).$$

*Suppose a further condition* $\sup_{j\in\mathbb{N}}\gamma_j^{-1}\|\Lambda^{1/2}\phi_j\|^2 < \infty$. *Then, as $n \to \infty$, if*

$$h_n^{-1} + g_n^{-1} + n^{-1/2}h_n^{3/2}(\log h_n)g_n^2(\log g_n) + n^{-1/2}g_n^{1/2}h_n^3(\log h_n)^2 \to 0,$$

*then for each $\eta > 0$,*

$$\mathsf{P}^*\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle(\hat{\Pi}_{h_n}^* - \Pi_{h_n})(\hat{\beta}_{g_n} - \beta), X_0\rangle| > \eta\Big|X_0\right) \xrightarrow{\mathsf{P}} 0.$$

*Proof.* Following the spirit of Lemma 17 and Remark 15, we ignore the remainder terms related to either $\mathcal{E}_j^c$, $\mathcal{A}_{h_n}^c$, $(\mathcal{E}_j^*)^c$, or $(\mathcal{A}_{h_n}^*)^c$. Based on Lemmas 20-22 and 31, and the decomposition

(3.38), one can see that

$$\begin{aligned}
&\mathsf{E}^*\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle(\hat{\Pi}^*_{h_n}-\Pi_{h_n})(\hat{\beta}_{g_n}-\Pi_{h_n}\beta),X_0\rangle|\Big|X_0\right]\\
=&\mathsf{E}^*\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle(\hat{\beta}_{g_n}-\Pi_{h_n}\beta),(\hat{\Pi}^*_{h_n}-\Pi_{h_n})X_0\rangle|\Big|X_0\right]\\
\leq&\mathsf{E}^*\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}\|(\hat{\beta}_{g_n}-\Pi_{h_n}\beta)\|\|(\hat{\Pi}^*_{h_n}-\Pi_{h_n})X_0\|\Big|X_0\right]\\
=&\sqrt{\frac{n}{s_{h_n}(X_0)}}\|(\hat{\beta}_{g_n}-\Pi_{h_n}\beta)\|\mathsf{E}^*[\|(\hat{\Pi}^*_{h_n}-\Pi_{h_n})X_0\|\Big|X_0]\\
=&O_{\mathsf{P}}\left(M_{n,g_n}h_n^{-1/2}\sum_{j=1}^{h_n}j\log j\right).
\end{aligned}$$

Meanwhile, as seen in the proof of Proposition 23, we have

$$\begin{aligned}
\hat{\Pi}^*_{h_n}-\Pi_{h_n}&=\frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}\left\{(zI-\Gamma^*_n)^{-1}-(zI-\Gamma)^{-1}\right\}dz\\
&=\mathcal{S}^*_n+\mathcal{R}^*_n+r^*_{1n}\mathbb{I}_{(\mathcal{A}^*_{h_n})^c}
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{S}^*_n&=\frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}(zI-\Gamma)^{-1}(\Gamma^*_n-\Gamma)(zI-\Gamma)^{-1}dz,\\
\mathcal{R}^*_n&=\frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}(zI-\Gamma)^{-1}(\Gamma^*_n-\Gamma)(zI-\Gamma)^{-1}(\Gamma^*_n-\Gamma)(zI-\Gamma^*_n)^{-1}dz.
\end{aligned}$$

Following the proof of Lemma 35, as $n\to\infty$, one can show that

$$\begin{aligned}
\frac{n}{h_n}\mathsf{E}\left[E^*\left[\langle\mathcal{S}^*_n(I-\Pi_{g_n})\beta,X_0\rangle^2|X_0\right]\right]&\leq Ch_n^{-1}\sum_{l=1}^{h_n}\gamma_l\left(\sum_{l'>g_n}|\beta_{l'}|\frac{\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_l-\gamma_{l'}}\right)^2\\
&\leq Ch_n^{-1}\sum_{l=1}^{h_n}\gamma_l\left(\sum_{l'>h_n}|\beta_{l'}|\frac{\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_l-\gamma_{l'}}\right)^2\to 0,
\end{aligned}$$

which implies that

$$\mathsf{E}^*\left[\frac{n}{s_{h_n}(X_0)}\langle\mathcal{S}^*_n(I-\Pi_{g_n})\beta,X_0\rangle^2\Big|X_0\right]=o_{\mathsf{P}}(1).$$

Next, note that $\|(I-\Pi_{g_n})\beta\| \leq \sqrt{\sum_{j>g_n}\beta_j^2}$. Finally, following the proof of Lemma 36, as $n\to\infty$, we have that

$$\mathsf{E}^*\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle\mathcal{R}_n^*(I-\Pi_{g_n})\beta, X_0\rangle| > \eta\Big|X_0\right] = O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sqrt{\sum_{j>g_n}\beta_j^2}\sum_{j=1}^{h_n}(j\log j)^2\right),$$

which completes the proof. □

#### 3.10.4 Variance term: lemmas for Proposition 16 in the main text

In what follows, we suppose that Conditions (A1)-(A7) and $\|\hat{\beta}_{g_n}-\beta\| \xrightarrow{\mathsf{P}} 0$ as $n\to\infty$. Recall $Z_{i,n}^* = \langle X_i^*\varepsilon_{i,g_n}^* - \tilde{U}_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle$ with $\tilde{U}_{n,g_n} \equiv n^{-1}\sum_{i=1}^n X_i^*\varepsilon_{i,g_n}^*$ and $\hat{v}_n^2 \equiv \sum_{i=1}^n \mathsf{E}^*[Z_{i,n}^{*2}|X_0]$ from the proof of Proposition 16.

**Lemma 37.** *As $n\to\infty$, we have that $n^{-1}\hat{v}_n^2 \sim_{\mathsf{P}} s_{h_n}(X_0)$ in the sense that*

$$\left|\frac{n^{-1}\hat{v}_n^2}{s_{h_n}(X_0)} - 1\right| \xrightarrow{\mathsf{P}} 0.$$

*Proof.* We first see that

$$\begin{aligned}\mathsf{E}^*[Z_{i,n}^{*2}|X_0] &= \mathsf{E}^*[\langle X_i^*\varepsilon_{i,g_n}^* - \mathsf{E}^*[X_i^*\varepsilon_{i,g_n}^*], \Gamma_{h_n}^{-1}X_0\rangle^2|X_0] \\ &= \langle \mathsf{E}^*[(X_i^*\varepsilon_{i,g_n}^* - \mathsf{E}^*[X_i^*\varepsilon_{i,g_n}^*])^{\otimes 2}]\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle.\end{aligned}$$

with $\mathsf{E}^*[(X_i^*\varepsilon_{i,g_n}^* - \mathsf{E}^*[X_i^*\varepsilon_{i,g_n}^*])^{\otimes 2}] = \mathsf{E}^*[(X_i^*\varepsilon_{i,g_n}^*)^{\otimes 2}] - (\mathsf{E}^*[X_i^*\varepsilon_{i,g_n}^*])^{\otimes 2}$. We then find that

$$\mathsf{E}^*[(X_i^*\varepsilon_{i,g_n}^* - \mathsf{E}^*[X_i^*\varepsilon_{i,g_n}^*])^{\otimes 2}] = \hat{\Lambda}_{n,g_n}.$$

This implies that

$$\begin{aligned}\mathsf{E}^*[Z_{i,n}^{*2}|X_0] &= \langle \hat{\Lambda}_{n,g_n}\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle \\ &= \langle (\hat{\Lambda}_{n,g_n} - \Lambda)\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle + s_{h_n}(X_0),\end{aligned}$$

and hence,

$$\left|\frac{n^{-1}\hat{v}_n^2}{s_{h_n}(X_0)} - 1\right| = s_{h_n}(X_0)^{-1}|\langle (\hat{\Lambda}_{n,g_n} - \Lambda)\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle|.$$

The result now follows from the fourth part of the proof of Proposition 20. □

**Lemma 38.** *As $n \to \infty$, if $n^{-1}h_n^2 \to 0$ and $\left(n^{-1}\sum_{j=1}^{h_n}\lambda_j^{-1}\right)\|\hat\beta_{g_n}-\beta\|^2 \xrightarrow{\mathsf{P}} 0$, then we have that*

$$\mathsf{E}^*\left[\left(\hat v_n^{-1}\max_{1\le i\le n}|Z_{i,n}^*|\right)^4\Big|X_0\right] \xrightarrow{\mathsf{P}} 0.$$

*Proof.* We first see that $|Z_{i,n}^*| \le |\langle X_i^*\varepsilon_{i,g_n}^*, \Gamma_{h_n}^{-1}X_0\rangle| + |\langle(\overline{X\hat\varepsilon})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle|$. By the second part of Lemma 23, the second term is given as

$$\hat v_n^{-1}|\langle(\overline{X\hat\varepsilon})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle| = n^{-1/2}s_{h_n}(X_0)^{-1/2}|\langle(\overline{X\hat\varepsilon})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle|\sqrt{\frac{s_{h_n}(X_0)}{n^{-1}v_h^2}} = o_{\mathsf{P}}(n^{-1/2}).$$

Note that

$$\hat v_n^{-1}\max_{1\le i\le n}|\langle X_i^*\varepsilon_{i,g_n}^*, \Gamma_{h_n}^{-1}X_0\rangle| \le n^{-1/2}\max_{1\le i\le n}\|\Lambda_{h_n}^{-1/2}X_i^*\varepsilon_{i,g_n}^*\|\left(\frac{\|\Lambda_{h_n}^{1/2}\Gamma_{h_n}^{-1}X_0\|^2}{n^{-1}\hat v_n^2}\right)^{1/2},$$

where

$$\frac{\|\Lambda_{h_n}^{1/2}\Gamma_{h_n}^{-1}X_0\|^2}{n^{-1}\hat v_n^2} \le \frac{s_{h_n}(X_0)}{n^{-1}\hat v_n^2} = 1 + o_{\mathsf{P}}(1) = O_{\mathsf{P}}(1)$$

as $n\to\infty$, since $\|\Lambda_{h_n}^{1/2}\Gamma_{h_n}^{-1}X_0\|^2 = \langle\Lambda_{h_n}\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle \le \langle\Lambda\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle = s_{h_n}(X_0)$.

We now need to deal with the term $n^{-1/2}\max_{1\le i\le n}\|\Lambda_{h_n}^{-1/2}X_i^*\varepsilon_{i,g_n}^*\|$. Inspired by the identity

$$X_i\hat\varepsilon_{i,g_n} = X_i\varepsilon_i - X_i^{\otimes 2}(\hat\beta_{g_n}-\beta),$$

we have that

$$\|\Lambda_{h_n}^{-1/2}X_i^*\varepsilon_{i,g_n}^*\|^2 \le 2\|\Lambda_{h_n}^{-1/2}\{X_i^*\varepsilon_{i,g_n}^* + (X_i^*)^{\otimes 2}(\hat\beta_{g_n}-\beta)\}\|^2 + 2\|\Lambda_{h_n}^{-1/2}(X_i^*)^{\otimes 2}(\hat\beta_{g_n}-\beta)\|^2.$$

To bound the first term in the above display, note that

$$\begin{aligned}\|\Lambda_{h_n}^{-1/2}\{X_i^*\varepsilon_{i,g_n}^* + (X_i^*)^{\otimes 2}(\hat\beta_{g_n}-\beta)\}\|^4 &= \left[\sum_{j=1}^{h_n}\lambda_j^{-1}\langle\{X_i^*\varepsilon_{i,g_n}^* + (X_i^*)^{\otimes 2}(\hat\beta_{g_n}-\beta)\}, \psi_j\rangle^2\right]^2\\ &\le h_n\sum_{j=1}^{h_n}\lambda_j^{-2}\langle\{X_i^*\varepsilon_{i,g_n}^* + (X_i^*)^{\otimes 2}(\hat\beta_{g_n}-\beta)\}, \psi_j\rangle^4\end{aligned}$$

by Jensen's inequality. By taking the maximum and the bootstrap expectation, we have

$$
\begin{aligned}
&n^{-2}\mathsf{E}^*\left[\max_{1\le i\le n}\|\Lambda_{h_n}^{-1/2}\{X_i^*\varepsilon_{i,g_n}^* + (X_i^*)^{\otimes 2}(\hat{\beta}_{g_n}-\beta)\}\|^4\right]\\
\le& n^{-2}h_n\sum_{i=1}^{n}\sum_{j=1}^{h_n}\lambda_j^{-2}\mathsf{E}^*[\langle\{X_i^*\varepsilon_{i,g_n}^* + (X_i^*)^{\otimes 2}(\hat{\beta}_{g_n}-\beta)\},\psi_j\rangle^4]\\
=& n^{-2}h_n\sum_{i=1}^{n}\sum_{j=1}^{h_n}\lambda_j^{-2}\left(n^{-1}\sum_{i'=1}^{n}\langle X_{i'}\varepsilon_{i'},\psi_j\rangle^4\right)\\
=& n^{-2}h_n\sum_{i=1}^{n}\sum_{j=1}^{h_n}\lambda_j^{-2}\langle X_i\varepsilon_i,\psi_j\rangle^4.
\end{aligned}
$$

From Condition (A7), we derive that

$$
\mathsf{E}\left[\mathsf{E}^*\left[\left[n^{-1/2}\max_{1\le i\le n}\|\Lambda_{h_n}^{-1/2}\{X_i^*\varepsilon_{i,g_n}^* + (X_i^*)^{\otimes 2}(\hat{\beta}_{g_n}-\beta)\}\|\right]^4\right]\right] \le Cn^{-1}h_n^2.
$$

For the next term, note that

$$
\|\Lambda_{h_n}^{-1/2}(X_i^*)^{\otimes 2}(\hat{\beta}_{g_n}-\beta)\|^2 = \sum_{j=1}^{h_n}\lambda_j^{-1}\langle(X_i^*)^{\otimes 2}(\hat{\beta}_{g_n}-\beta),\psi_j\rangle^2,
$$

and hence,

$$
\begin{aligned}
&\mathsf{E}^*[n^{-1}\|\Lambda_{h_n}^{-1/2}(X_i^*)^{\otimes 2}(\hat{\beta}_{g_n}-\beta)\|^2]\\
=& n^{-1}\sum_{j=1}^{h_n}\lambda_j^{-1}\langle\mathsf{E}^*[(X_i^*)^{\otimes 2}](\hat{\beta}_{g_n}-\beta),\psi_j\rangle^2 = n^{-1}\sum_{j=1}^{h_n}\lambda_j^{-1}\langle\tilde{\Gamma}_n(\hat{\beta}_{g_n}-\beta),\psi_j\rangle^2\\
=& O_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{h_n}\lambda_j^{-1}\|\hat{\beta}_{g_n}-\beta\|^2\right),
\end{aligned}
$$

where $\tilde{\Gamma}_n \equiv n^{-1}\sum_{i=1}^{n}X_i^{\otimes 2}$, since $\mathsf{E}[\|\tilde{\Gamma}_n-\Gamma\|_\infty^2] \le n^{-1}\mathsf{E}[\|X_1\|^4]$ from Theorem 2.5 in [32].

In summary, we have that

$$
\mathsf{E}^*\left[\left(\hat{v}_n^{-1}\max_{1\le i\le n}|Z_{i,n}^*|\right)^4\middle|X_0\right] = O_{\mathsf{P}}(n^{-1}h_n^2) + O_{\mathsf{P}}\left(\left\{\left(n^{-1}\sum_{j=1}^{h_n}\lambda_j^{-1}\right)\|\hat{\beta}_{g_n}-\beta\|^2\right\}^2\right) + o_{\mathsf{P}}(n^{-2}).
$$

□

**Lemma 39.** *As $n\to\infty$, we have that*

$$
\mathsf{E}^*\left[\left|\frac{n^{-1}\sum_{i=1}^{n}Z_{i,n}^{*2}}{s_{h_n}(X_0)}-1\right|\middle|X_0\right] \xrightarrow{\mathsf{P}} 0.
$$

*Proof.* Note that

$$n^{-1}\sum_{i=1}^{n} Z_{i,n}^{*2} = \langle \hat{\Lambda}_n^* \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0\rangle = \langle (\hat{\Lambda}_n^* - \Lambda)\Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0\rangle + s_{h_n}(X_0)$$

where $\check{\Lambda}_n^* \equiv n^{-1}\sum_{i=1}^{n}(X_i^*\varepsilon_{i,g_n}^* - (\overline{X\hat{\varepsilon}})_{n,g_n})^{\otimes 2}$ with its mean $\mathsf{E}^*[\check{\Lambda}_n^*] = \hat{\Lambda}_{n,g_n}$. Here, $(\overline{X\hat{\varepsilon}})_{n,g_n}$ is defined in (3.45). Inspired by the identity

$$X_i\hat{\varepsilon}_{i,g_n} = X_i\varepsilon_i - X_i^{\otimes 2}(\hat{\beta}_{g_n} - \beta),$$

we see that

$$\mathsf{E}^*[f(X_i^*\varepsilon_{i,g_n}^* + (X_i^*)^{\otimes 2}(\hat{\beta}_{g_n} - \beta))] = n^{-1}\sum_{i=1}^{n} f(X_i\varepsilon_i) \tag{3.65}$$

for any function $f : \mathbb{H} \to \mathbb{R}$. From the following decomposition

$$X_i^*\varepsilon_{i,g_n}^* - (\overline{X\hat{\varepsilon}})_{n,g_n} = X_i^*\varepsilon_{i,g_n}^* + (X_i^*)^{\otimes 2}(\hat{\beta}_{g_n} - \beta) - (X_i^*)^{\otimes 2}(\hat{\beta}_{g_n} - \beta) - (\overline{X\hat{\varepsilon}})_{n,g_n},$$

we have that

$$\begin{aligned}
&\langle (\hat{\Lambda}_n^* - \Lambda)\Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0\rangle \\
=& n^{-1}\sum_{i=1}^{n}\langle [\{X_i^*\varepsilon_{i,g_n}^* + (X_i^*)^{\otimes 2}(\hat{\beta}_{g_n} - \beta)\}^{\otimes 2} - \Lambda]\Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0\rangle \qquad (3.66)\\
&+ n^{-1}\sum_{i=1}^{n}\langle (X_i^*)^{\otimes 2}(\hat{\beta}_{g_n} - \beta), \Gamma_{h_n}^{-1} X_0\rangle^2 \\
&+ \langle (\overline{X\hat{\varepsilon}})_{n,g_n}, \Gamma_{h_n}^{-1} X_0\rangle^2 \\
&- 2n^{-1}\sum_{i=1}^{n}\langle X_i^*\varepsilon_{i,g_n}^*, \Gamma_{h_n}^{-1} X_0\rangle\langle (X_i^*)^{\otimes 2}(\hat{\beta}_{g_n} - \beta), \Gamma_{h_n}^{-1} X_0\rangle \\
&- 2n^{-1}\sum_{i=1}^{n}\langle X_i^*\varepsilon_{i,g_n}^*, \Gamma_{h_n}^{-1} X_0\rangle\langle (\overline{X\hat{\varepsilon}})_{n,g_n}, \Gamma_{h_n}^{-1} X_0\rangle \\
&+ 2n^{-1}\sum_{i=1}^{n}\langle (X_i^*)^{\otimes 2}(\hat{\beta}_{g_n} - \beta), \Gamma_{h_n}^{-1} X_0\rangle\langle (\overline{X\hat{\varepsilon}})_{n,g_n}, \Gamma_{h_n}^{-1} X_0\rangle.
\end{aligned}$$

We now investigate an upper bound for each term in the preceding display.

The first term in (3.66) is bounded as follows. By (3.65), we have that

$$\begin{aligned}
&\mathsf{E}^*[|\langle[\{X_i^*\varepsilon_{i,g_n}^* + (X_i^*)^{\otimes 2}(\hat{\beta}_{g_n} - \beta)\}^{\otimes 2} - \Lambda]\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle||X_0] \\
=&n^{-1}\sum_{i=1}^{n}|\langle\{(X_i\varepsilon_i)^{\otimes 2} - \Lambda\}\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle| \\
\leq&n^{-1}\sum_{i=1}^{n}\|(X_i\varepsilon_i)^{\otimes 2} - \Lambda\|\|\Gamma_{h_n}^{-1}X_0\|^2
\end{aligned}$$

Recall that $\mathsf{E}[\|(X_i\varepsilon_i)^{\otimes 2} - \Lambda\|^2] \leq Cn^{-1}$ from Theorem 2.3 of [32] and $\mathsf{E}[\|\Gamma_{h_n}^{-1}X_0\|^2] = \sum_{j=1}^{h_n}\gamma_j^{-1}$. This implies that

$$\begin{aligned}
&\mathsf{E}\left[\mathsf{E}^*\left[\left|n^{-1}\sum_{i=1}^{n}\langle[\{X_i^*\varepsilon_{i,g_n}^* + (X_i^*)^{\otimes 2}(\hat{\beta}_{g_n} - \beta)\}^{\otimes 2} - \Lambda]\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle\right|\middle|X_0\right]\right] \\
\leq&Cn^{-1/2}\sum_{j=1}^{h_n}\gamma_j^{-1},
\end{aligned}$$

which implies that

$$\begin{aligned}
&\mathsf{E}^*\left[s_{h_n}(X_0)^{-1}\left|n^{-1}\sum_{i=1}^{n}\langle[\{X_i^*\varepsilon_{i,g_n}^* + (X_i^*)^{\otimes 2}(\hat{\beta}_{g_n} - \beta)\}^{\otimes 2} - \Lambda]\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle\right|\middle|X_0\right] \\
=&O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right),
\end{aligned}$$

where this convereges to zero under Condition (A5).

The second term in (3.66) is bounded as follows. Since

$$\begin{aligned}
L_n \equiv& \mathsf{E}^*[\langle(X_i^*)^{\otimes 2}(\hat{\beta}_{g_n} - \beta), \Gamma_{h_n}^{-1}X_0\rangle^2|X_0] \\
=& n^{-1}\sum_{i=1}^{n}\langle X_i^{\otimes 2}(\hat{\beta}_{g_n} - \beta), \Gamma_{h_n}^{-1}X_0\rangle^2 \\
=& n^{-1}\sum_{i=1}^{n}\langle X_i, \hat{\beta}_{g_n} - \beta\rangle^2\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2 \\
=& \left\langle n^{-1}\sum_{i=1}^{n}\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2 X_i^{\otimes 2}(\hat{\beta}_{g_n} - \beta), \hat{\beta}_{g_n} - \beta\right\rangle,
\end{aligned}$$

by the third part of Lemma 24, we have that $s_{h_n}(X_0)^{-1}L_n = O_{\mathsf{P}}(\|\hat{\beta}_{g_n} - \beta\|^2)$.

The third term in (3.66) is bounded as

$$s_{h_n}(X_0)^{-1}\langle(\overline{X\hat{\varepsilon}})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle^2 = O_{\mathsf{P}}\left(n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right) + O_{\mathsf{P}}(\|\hat{\beta}_{g_n} - \beta\|^2)$$

by the second part of Lemma 23.

The fourth term in (3.66) is bounded as follows. Note that

$$\begin{aligned}
L_n &\equiv \mathsf{E}^*[|\langle X_i^*\varepsilon_{i,g_n}^*, \Gamma_{h_n}^{-1}X_0\rangle\langle (X_i^*)^{\otimes 2}(\hat{\beta}_{g_n}-\beta), \Gamma_{h_n}^{-1}X_0\rangle||X_0] \\
&= n^{-1}\sum_{i=1}^n |\langle X_i\hat{\varepsilon}_{i,g_n}, \Gamma_{h_n}^{-1}X_0\rangle\langle X_i^{\otimes 2}(\hat{\beta}_{g_n}-\beta), \Gamma_{h_n}^{-1}X_0\rangle| \\
&\leq n^{-1}\sum_{i=1}^n |\langle X_i\varepsilon_i, \Gamma_{h_n}^{-1}X_0\rangle\langle X_i^{\otimes 2}(\hat{\beta}_{g_n}-\beta), \Gamma_{h_n}^{-1}X_0\rangle| \\
&+ n^{-1}\sum_{i=1}^n \langle X_i^{\otimes 2}(\hat{\beta}_{g_n}-\beta), \Gamma_{h_n}^{-1}X_0\rangle^2 \\
&= n^{-1}\sum_{i=1}^n |\langle\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2 X_i\varepsilon_i, \hat{\beta}_{g_n}-\beta\rangle| \\
&+ \left\langle n^{-1}\sum_{i=1}^n \langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2 X_i^{\otimes 2}(\hat{\beta}_{g_n}-\beta), \hat{\beta}_{g_n}-\beta\right\rangle \\
&\leq \left(n^{-1}\sum_{i=1}^n \langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2\|X_i\varepsilon_i\|\right)\|\hat{\beta}_{g_n}-\beta\| \\
&+ \left(n^{-1}\sum_{i=1}^n \|X_i\|^2\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2\right)\|\hat{\beta}_{g_n}-\beta\|^2
\end{aligned}$$

By the third parts of each of Lemmas 24-25, we have that $s_{h_n}(X_0)^{-1}L_n = O_{\mathsf{P}}(\|\hat{\beta}_{g_n}-\beta\|)$.

The fifth term in (3.66) is bounded as follows. Note that

$$\begin{aligned}
L_n &\equiv \mathsf{E}^*[|\langle X_i^*\varepsilon_{i,g_n}^*, \Gamma_{h_n}^{-1}X_0\rangle\langle(\overline{X\hat{\varepsilon}})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle||X_0] \\
&= n^{-1}\sum_{i=1}^n |\langle X_i\hat{\varepsilon}_{i,g_n}, \Gamma_{h_n}^{-1}X_0\rangle\langle(\overline{X\hat{\varepsilon}})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle| \\
&\leq |\langle(\overline{X\hat{\varepsilon}})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle| \left\{n^{-1}\sum_{i=1}^n |\langle X_i\varepsilon_i, \Gamma_{h_n}^{-1}X_0\rangle| + n^{-1}\sum_{i=1}^n |\langle X_i^{\otimes 2}(\hat{\beta}_{g_n}-\beta), \Gamma_{h_n}^{-1}X_0\rangle|\right\}
\end{aligned}$$

We observe that

$$n^{-1}\sum_{i=1}^n |\langle X_i\varepsilon_i, \Gamma_{h_n}^{-1}X_0\rangle| \leq \left(n^{-1}\sum_{i=1}^n \langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2\right)^{1/2}\left(n^{-1}\sum_{i=1}^n \varepsilon_i^2\right)^{1/2}$$

since $n^{-1}\sum_{i=1}^n \|X_i\varepsilon_i\| \xrightarrow{\mathsf{P}} \mathsf{E}[\|X\varepsilon\|] < \infty$ and

$$\begin{aligned}
&n^{-1}\sum_{i=1}^n |\langle X_i^{\otimes 2}(\hat\beta_{g_n}-\beta), \Gamma_{h_n}^{-1}X_0\rangle| \\
=&n^{-1}\sum_{i=1}^n |\langle\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle X_i, (\hat\beta_{g_n}-\beta)\rangle| \le \left(n^{-1}\sum_{i=1}^n |\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle|\|X_i\|\right)\|\hat\beta_{g_n}-\beta\| \\
\le&\left(n^{-1}\sum_{i=1}^n \langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2\right)^{1/2}\left(n^{-1}\sum_{i=1}^n \|X_i\|^2\right)^{1/2}\|\hat\beta_{g_n}-\beta\|.
\end{aligned}$$

Here, $\mathsf{E}[\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2] = \sum_{j=1}^{h_n}\gamma_j^{-1}\mathsf{E}[\langle X_i, \phi_j\rangle^2] = h_n$ since the FPC scores $\xi_j$ are uncorrelated random variables with mean zero and variance $\gamma_j$. This implies that $\mathsf{E}\left[n^{-1}\sum_{i=1}^n \langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2\right] = h_n$. Therefore, we have

$$n^{-1}\sum_{i=1}^n |\langle X_i\varepsilon_i, \Gamma_{h_n}^{-1}X_0\rangle| + n^{-1}\sum_{i=1}^n |\langle X_i^{\otimes 2}(\hat\beta_{g_n}-\beta), \Gamma_{h_n}^{-1}X_0\rangle| = O_{\mathsf{P}}(h_n^{1/2}).$$

Since

$$s_{h_n}(X_0)^{-1/2}|\langle(\overline{X\hat\varepsilon})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle| = O_{\mathsf{P}}\left(\left(n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2}\right) + O_{\mathsf{P}}(\|\hat\beta_{g_n}-\beta\|)$$

from the second part of Lemma 23, we conclude that

$$s_{h_n}(X_0)^{-1}L_n = O_{\mathsf{P}}\left(\left(n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2}\right) + O_{\mathsf{P}}(\|\hat\beta_{g_n}-\beta\|).$$

The sixth term in (3.66) is bounded as follows. Note that

$$\begin{aligned}
L_n &\equiv \mathsf{E}^*[|\langle (X_i^*)^{\otimes 2}(\hat\beta_{g_n}-\beta), \Gamma_{h_n}^{-1}X_0\rangle\langle(\overline{X\hat\varepsilon})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle||X_0] \\
&= n^{-1}\sum_{i=1}^n |\langle X_i^{\otimes 2}(\hat\beta_{g_n}-\beta), \Gamma_{h_n}^{-1}X_0\rangle\langle(\overline{X\hat\varepsilon})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle| \\
&= \left(n^{-1}\sum_{i=1}^n |\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle|\|X_i\|\right)\|\hat\beta_{g_n}-\beta\||\langle(\overline{X\hat\varepsilon})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle| \\
&\le \left(n^{-1}\sum_{i=1}^n \langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2\right)^{1/2}\left(n^{-1}\sum_{i=1}^n \|X_i\|^2\right)^{1/2}\|\hat\beta_{g_n}-\beta\||\langle(\overline{X\hat\varepsilon})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle|.
\end{aligned}$$

As the previous paragraph, we conclude that

$$s_{h_n}(X_0)^{-1}L_n = O_{\mathsf{P}}\left(\left(n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2}\|\hat\beta_{g_n}-\beta\|\right) + O_{\mathsf{P}}(\|\hat\beta_{g_n}-\beta\|^2).$$

In summary, we have that

$$\begin{aligned}&\mathsf{E}^*\left[\left|\frac{n^{-1}\sum_{i=1}^n (Z^*_{i,n})^2}{s_{h_n}(X_0)}-1\right|\Big|X_0\right]\\ =&O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)+O_{\mathsf{P}}\left(n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)+O_{\mathsf{P}}(\|\hat{\beta}_{g_n}-\beta\|).\end{aligned}$$

Note that Condition (A5) implies that $n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\le n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\to 0$ as $n\to\infty$. Thus, as $n\to\infty$, if $\|\hat{\beta}_{g_n}-\beta\|\to 0$, we have that

$$\mathsf{E}^*\left[\left|\frac{n^{-1}\sum_{i=1}^n (Z^*_{i,n})^2}{s_{h_n}(X_0)}-1\right|\Big|X_0\right]\xrightarrow{\mathsf{P}}0.$$

□

**Lemma 40.** *As $n\to\infty$, we have that*

$$\mathsf{E}^*\left[\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\bar{X}^*(\overline{\varepsilon^*})_{n,g_n}-\bar{X}(\bar{\hat{\varepsilon}})_{g_n},\Gamma_{h_n}^{-1}X_0\rangle\right)^2\Big|X_0\right]\xrightarrow{\mathsf{P}}0.$$

*Proof.* From Lemma 30, we have that

$$\begin{aligned}&\mathsf{E}^*\left[\frac{n}{s_{h_n}(X_0)}\langle\bar{X}^*(\overline{\varepsilon^*})_{n,g_n}-\bar{X}(\bar{\hat{\varepsilon}})_{g_n},\Gamma_{h_n}^{-1}X_0\rangle^2\Big|X_0\right]\\ \le&\{h_ns_{h_n}(X_0)^{-1}\}(n\mathsf{E}^*[\|\bar{X}^*(\overline{\varepsilon^*})_{n,g_n}-\bar{X}(\bar{\hat{\varepsilon}})_{g_n}\|^2])(h_n^{-1}\|\Gamma_{h_n}^{-1}X_0\|^2)\\ =&O_{\mathsf{P}}(1)O_{\mathsf{P}}(n^{-1})O_{\mathsf{P}}\left(h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)\\ =&O_{\mathsf{P}}\left(n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right),\end{aligned}$$

where the last big $O_{\mathsf{P}}$ term converges to zero under Condition (A5). □

### 3.10.5 Scaling term

We investigate the consistency of the ratio of $\hat{s}^*_{k_n,h_n,g_n}(X_0)$ to $\hat{s}_{h_n}(X_0)$ (or to $s_{h_n}(X_0)$) to 1 in the bootstrap probability $\mathsf{P}^*$. The bootstrap scaling $\hat{s}^*_{h_n}(X_0)$ can be decomposed in a similar way to the decomposition (3.43) as follows:

$$
\begin{aligned}
&\hat{s}^*_{h_n}(X_0) = \langle \hat{\Lambda}^*_{n,k_n,g_n}(\hat{\Gamma}^*_{h_n})^{-1}X_0, (\hat{\Gamma}^*_{h_n})^{-1}X_0\rangle \\
=&\langle \hat{\Lambda}_{n,k_n}\{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}X_0, (\hat{\Gamma}^*_{h_n})^{-1}X_0\rangle + \langle \hat{\Lambda}^*_{n,k_n,g_n}\Gamma^{-1}_{h_n}X_0, (\hat{\Gamma}^*_{h_n})^{-1}X_0\rangle \\
=&\langle \hat{\Lambda}^*_{n,k_n,g_n}\{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}X_0, \{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}X_0\rangle + \langle \hat{\Lambda}^*_{n,k_n,g_n}\{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}X_0, \Gamma^{-1}_{h_n}X_0\rangle \\
&+ \langle \hat{\Lambda}^*_{n,k_n,g_n}\Gamma^{-1}_{h_n}X_0, \{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}X_0\rangle + \langle \hat{\Lambda}^*_{n,k_n,g_n}\Gamma^{-1}_{h_n}X_0, \Gamma^{-1}_{h_n}X_0\rangle \\
=&\langle \hat{\Lambda}^*_{n,k_n,g_n}\{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}X_0, \{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}X_0\rangle \\
&+ 2\langle \hat{\Lambda}^*_{n,k_n,g_n}\Gamma^{-1}_{h_n}X_0, \{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}X_0\rangle + \langle \hat{\Lambda}^*_{n,k_n,g_n}\Gamma^{-1}_{h_n}X_0, \Gamma^{-1}_{h_n}X_0\rangle \\
=&\langle (\hat{\Lambda}^*_{n,k_n,g_n} - \Lambda)\{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}X_0, \{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}X_0\rangle \qquad (3.67)\\
&+ \langle \Lambda\{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}X_0, \{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}X_0\rangle \\
&+ 2\langle (\hat{\Lambda}^*_{n,k_n,g_n} - \Lambda)\Gamma^{-1}_{h_n}X_0, \{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}X_0\rangle \\
&+ 2\langle \Lambda\Gamma^{-1}_{h_n}X_0, \{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}X_0\rangle \\
&+ \langle (\hat{\Lambda}^*_{n,k_n,g_n} - \Lambda)\Gamma^{-1}_{h_n}X_0, \Gamma^{-1}_{h_n}X_0\rangle + \langle \Lambda\Gamma^{-1}_{h_n}X_0, \Gamma^{-1}_{h_n}X_0\rangle.
\end{aligned}
$$

Also, with $\varepsilon^*_{i,g_n} = Y^*_i - \langle \hat{\beta}_{g_n}, X^*_i\rangle$, we see by putting $\hat{\varepsilon}^*_{i,k_n,g_n} = \varepsilon^*_{i,g_n} - \langle \hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}, X^*_i\rangle$ that

$$
\begin{aligned}
\hat{\Lambda}^*_{n,k_n,g_n} - \Lambda =& n^{-1}\sum_{i=1}^{n}(X^*_i\hat{\varepsilon}^*_{i,k_n,g_n})^{\otimes 2} - (\overline{X^*\hat{\varepsilon}^*})^{\otimes 2}_{n,k_n,g_n} - \Lambda \\
=& \tilde{\Lambda}^*_{n,g_n} - \Lambda + n^{-1}\sum_{i=1}^{n}(X^*_i)^{\otimes 2}\langle \hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}, X^*_i\rangle^2 \qquad (3.68)\\
&- 2n^{-1}\sum_{i=1}^{n}(X^*_i\varepsilon^*_{i,g_n}\otimes X^*_i)\langle \hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}, X^*_i\rangle - (\overline{X^*\hat{\varepsilon}^*})^{\otimes 2}_{n,k_n,g_n}
\end{aligned}
$$

where $\tilde{\Lambda}^*_{n,g_n} \equiv n^{-1}\sum_{i=1}^{n}(X^*_i\varepsilon^*_{i,g_n})^{\otimes 2}$ and

$$
\begin{aligned}
(\overline{X^*\hat{\varepsilon}^*})_{n,k_n,g_n} \equiv& n^{-1}\sum_{i=1}^{n}X^*_i\hat{\varepsilon}^*_{i,k_n,g_n} = n^{-1}\sum_{i=1}^{n}X^*_i\varepsilon^*_{i,g_n} - n^{-1}\sum_{i=1}^{n}\langle \hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}, X^*_i\rangle X^*_i \\
=& (\overline{X^*\varepsilon^*})_{n,g_n} - (\tilde{\Gamma}^*_n - \Gamma)(\hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}) - \Gamma(\hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}) \qquad (3.69)
\end{aligned}
$$

with $(\overline{X^*\varepsilon^*})_{n,g_n} \equiv n^{-1}\sum_{i=1}^{n}X^*_i\varepsilon^*_{i,g_n}$.

In what follows, we suppose that Conditions (A1)-(A8) hold and $\|\hat{\beta}_{g_n} - \beta\| \xrightarrow{\mathsf{P}} 0$ and for each $\eta > 0$, $\mathsf{P}^*(\|\hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}\| > \eta) \xrightarrow{\mathsf{P}} 0$ as $n \to \infty$.

**Lemma 41.** *Define* $\check{\Lambda}^*_n \equiv n^{-1}\sum_{i=1}^n (X_i^*Y_i^* - X_i^*\langle X_i^*, \beta\rangle)^{\otimes 2}$. *Then, we have* $\mathsf{E}[\mathsf{E}^*[n\|\check{\Lambda}^*_n - \Lambda\|_{HS}^2]] = O(1)$, *which implies that for each* $\eta > 0$

$$\mathsf{P}^*(\|\tilde{\Lambda}^*_{n,g_n} - \Lambda\|_\infty > \eta) \xrightarrow{\mathsf{P}} 0.$$

*Proof.* Note that

$$\begin{aligned}
\tilde{\Lambda}^*_{n,g_n} &\equiv n^{-1}\sum_{i=1}^n (X_i^*\varepsilon^*_{i,g_n})^{\otimes 2} = n^{-1}\sum_{i=1}^n (X_i^*Y_i^* - X_i^*\langle X_i^*, \beta\rangle - X_i^*\langle X_i^*, \hat{\beta}_{g_n} - \beta\rangle)^{\otimes 2}\\
&= \check{\Lambda}^*_n + n^{-1}\sum_{i=1}^n (X_i^*\langle X_i^*, \hat{\beta}_{g_n} - \beta\rangle)^{\otimes 2}\\
&\quad - n^{-1}\sum_{i=1}^n (X_i^*Y_i^* - X_i^*\langle X_i^*, \beta\rangle) \otimes (X_i^*\langle X_i^*, \hat{\beta}_{g_n} - \beta\rangle)\\
&\quad - n^{-1}\sum_{i=1}^n (X_i^*\langle X_i^*, \hat{\beta}_{g_n} - \beta\rangle) \otimes (X_i^*Y_i^* - X_i^*\langle X_i^*, \beta\rangle)
\end{aligned}$$

To see the convergence of the first term, write $L_i^* = (X_i^*Y_i^* - X_i^*\langle X_i^*, \beta\rangle)^{\otimes 2}$. Since $\{L_i^*\}_{i=1}^n$ are iid with mean $\mathsf{E}^*[L_i^*] = n^{-1}\sum_{i=1}^n (X_i\varepsilon_i)^{\otimes 2} \equiv \tilde{\Lambda}_n$ under the bootstrap probability $\mathsf{P}^*$, we have

$$\begin{aligned}
\mathsf{E}^*\left[\left\| n^{-1}\sum_{i=1}^n (L_i^* - \tilde{\Lambda}_n)\right\|_{HS}^2\right] &= n^{-1}\sum_{i=1}^n \mathsf{E}^*[\|L_i^* - \tilde{\Lambda}_n\|_{HS}^2] = n^{-2}\sum_{i=1}^n \|(X_i\varepsilon_i)^{\otimes 2} - \tilde{\Lambda}_n\|_{HS}^2\\
&\le C\left(n^{-2}\sum_{i=1}^n \|X_i\varepsilon_i\|^4 + n^{-1}\|\tilde{\Lambda}_n - \Lambda\|_{HS}^2 + n^{-1}\|\Lambda\|_{HS}^2\right),
\end{aligned}$$

which implies that

$$\begin{aligned}
\mathsf{E}[\mathsf{E}^*[n\|\check{\Lambda}^*_n - \Lambda\|_{HS}^2]] &= \mathsf{E}[\|X\varepsilon\|^4] + \mathsf{E}[n\|\tilde{\Lambda}_n - \Lambda\|_{HS}^2] + \mathsf{E}[\|\tilde{\Lambda}_n - \Lambda\|_{HS}^2] + \|\Lambda\|_{HS}^2\\
&= O(1)
\end{aligned}$$

For the rest of terms, note that

$$\begin{aligned}
&\mathsf{E}^*\left[\left\| n^{-1}\sum_{i=1}^n (X_i^*Y_i^* - X_i^*\langle X_i^*, \beta\rangle) \otimes (X_i^*\langle X_i^*, \hat{\beta}_{g_n} - \beta\rangle)\right\|_\infty\right]\\
\le& n^{-1}\sum_{i=1}^n \mathsf{E}^*\left[\|(X_i^*Y_i^* - X_i^*\langle X_i^*, \beta\rangle) \otimes (X_i^*\langle X_i^*, \hat{\beta}_{g_n} - \beta\rangle)\|_\infty\right]\\
=& n^{-1}\sum_{i=1}^n \|(X_iY_i - X_i\langle X_i, \beta\rangle) \otimes (X_i\langle X_i, \hat{\beta}_{g_n} - \beta\rangle)\|_\infty\\
\le& n^{-1}\sum_{i=1}^n \|X_i\varepsilon_i\|\|X_i\|^2\|\hat{\beta}_{g_n} - \beta\| \le \left(n^{-1}\sum_{i=1}^n \|X_i\varepsilon_i\|^2\right)\left(n^{-1}\sum_{i=1}^n \|X_i\|^4\right)\|\hat{\beta}_{g_n} - \beta\| \xrightarrow{\mathsf{P}} 0
\end{aligned}$$

and

$$\begin{aligned}
&\mathsf{E}^*\left[\left\|n^{-1}\sum_{i=1}^{n}(X_i^*\langle X_i^*,\hat{\beta}_{g_n}-\beta\rangle)^{\otimes 2}\right\|_\infty\right] \\
\le & n^{-1}\sum_{i=1}^{n}\mathsf{E}^*\left[\|(X_i^*\langle X_i^*,\hat{\beta}_{g_n}-\beta\rangle)^{\otimes 2}\|_\infty\right] = n^{-1}\sum_{i=1}^{n}\|(X_i\langle X_i,\hat{\beta}_{g_n}-\beta\rangle)^{\otimes 2}\|_\infty \\
\le & \left(n^{-1}\sum_{i=1}^{n}\|X_i\|^4\right)\|\hat{\beta}_{g_n}-\beta\|^2 \xrightarrow{\mathsf{P}} 0.
\end{aligned}$$

We thus have the desired result. □

**Lemma 42.**

1. *For each $\eta > 0$, we have $\mathsf{P}^*(\|(\overline{X^*\hat{\varepsilon}^*})_{n,k_n,g_n}\| > \eta) \xrightarrow{\mathsf{P}} 0$.*

2. *As $n \to \infty$, if $n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1} \to 0$ (which is implied by (A5)), for each $\eta > 0$, we have*

$$\mathsf{P}^*(s_{h_n}(X_0)^{-1}\langle(\overline{X^*\hat{\varepsilon}^*})_{n,k_n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle^2 > \eta | X_0) \xrightarrow{\mathsf{P}} 0.$$

*Proof.* Note that $\mathsf{E}^*[(\overline{X^*\varepsilon^*})_{n,g_n}] = \mathsf{E}^*[X_i^*\varepsilon_{i,g_n}^*] = (\overline{X\hat{\varepsilon}})_{n,g_n}$. Since $\{X_i^*\varepsilon_{i,g_n}^*\}_{i=1}^n$ are iid with mean $(\overline{X\hat{\varepsilon}})_{n,g_n}$, we have

$$\begin{aligned}
&\mathsf{E}^*\left[\|(\overline{X^*\varepsilon^*})_{n,g_n} - (\overline{X\hat{\varepsilon}})_{n,g_n}\|^2\right] \\
=& n^{-1}\mathsf{E}^*\left[\|X_i^*\varepsilon_{i,g_n}^* - (\overline{X\hat{\varepsilon}})_{n,g_n}\|^2\right] \le \frac{2}{n}\mathsf{E}^*[\|X_i^*\varepsilon_{i,g_n}^*\|^2] + \|(\overline{X\hat{\varepsilon}})_{n,g_n}\|^2 \\
=& \frac{2}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\|X_i\hat{\varepsilon}_{i,g_n}\|^2 + \|(\overline{X\hat{\varepsilon}})_{n,g_n}\|^2\right) \\
\le & \frac{2}{n}\left\{\left(\frac{2}{n}\sum_{i=1}^{n}\|X_i\varepsilon_i\|^2 + \frac{2}{n}\sum_{i=1}^{n}\|X_i\|^4\|\hat{\beta}_{g_n}-\beta\|^2\right) + \|(\overline{X\hat{\varepsilon}})_{n,g_n}\|^2\right\} \\
=& O_{\mathsf{P}}(n^{-1})
\end{aligned}$$

from the fact that $(\overline{X\hat{\varepsilon}})_{n,g_n} \xrightarrow{\mathsf{P}} 0$ by the first part of Lemma 23, which again implies that $\mathsf{E}^*[\|(\overline{X^*\varepsilon^*})_{n,g_n}\|^2] \xrightarrow{\mathsf{P}} 0$. Next, note that $\mathsf{E}^*[\tilde{\Gamma}_n^*] = \mathsf{E}^*[(X_i^*)^{\otimes 2}] = n^{-1}\sum_{i=1}^{n}X_i^{\otimes 2} = \tilde{\Gamma}_n$. Since

$\{(X_i^*)^{\otimes 2}\}_{i=1}^n$ are iid with mean $\tilde{\Gamma}_n$, we have

$$\begin{aligned}\mathsf{E}^*\left[\|\tilde{\Gamma}_n^* - \tilde{\Gamma}_n\|_{HS}^2\right] &= n^{-1}\mathsf{E}^*\left[\|(X_i^*)^{\otimes 2} - \tilde{\Gamma}_n\|_{HS}^2\right] \le \frac{2}{n}(\mathsf{E}^*[\|X_i^*\|^2] + \|\tilde{\Gamma}_n\|_{HS}^2) \\ &= \frac{2}{n}\left(n^{-1}\sum_{i=1}^n \|X_i\|^2 + 2\|\tilde{\Gamma}_n - \Gamma\|_{HS}^2 + 2\|\Gamma\|_{HS}^2\right) \\ &= O_{\mathsf{P}}(n^{-1})\end{aligned}$$

from the fact that $\mathsf{E}[\|\tilde{\Gamma}_n - \Gamma\|_{HS}^2] = O(n^{-1})$ by Theorem 2.5 of [32], which again implies that $\mathsf{E}^*[\|\tilde{\Gamma}_n^* - \Gamma\|_\infty^2] \le \mathsf{E}^*[\|\tilde{\Gamma}_n^* - \Gamma\|_{HS}^2] = O_{\mathsf{P}}(n^{-1})$. The first part then follows from the decomposition (3.69) and Theorem 10.

For the second part, note that

$$\begin{aligned}&s_{h_n}(X_0)^{-1}\langle(\overline{X^*\varepsilon^*})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle^2 \\ \le& 2s_{h_n}(X_0)^{-1}\langle(\overline{X^*\varepsilon^*})_{n,g_n} - (\overline{X\hat{\varepsilon}})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle^2 \\ &+ 2s_{h_n}(X_0)^{-1}\langle(\overline{X\hat{\varepsilon}})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle^2.\end{aligned}$$

Since

$$\begin{aligned}&\mathsf{E}^*[s_{h_n}(X_0)^{-1}\langle(\overline{X^*\varepsilon^*})_{n,g_n} - (\overline{X\hat{\varepsilon}})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle^2|X_0] \\ \le& s_{h_n}(X_0)^{-1}\mathsf{E}^*[\|(\overline{X^*\varepsilon^*})_{n,g_n} - (\overline{X\hat{\varepsilon}})_{n,g_n}\|^2]\|\Gamma_{h_n}^{-1}X_0\|^2 \\ =& O_{\mathsf{P}}\left(n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)\end{aligned}$$

from the fact that $\mathsf{E}^*[\|(\overline{X^*\varepsilon^*})_{n,g_n} - (\overline{X\hat{\varepsilon}})_{n,g_n}\|^2] = O_{\mathsf{P}}(1)$ as seen above and $s_{h_n}(X_0)^{-1}\langle(\overline{X\hat{\varepsilon}})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle^2 \xrightarrow{\mathsf{P}} 0$ by the second part of Lemma 23, we have $\mathsf{E}^*[s_{h_n}(X_0)^{-1}\langle(\overline{X^*\varepsilon^*})_{n,g_n}, \Gamma_{h_n}^{-1}X_0\rangle^2|X_0] \xrightarrow{\mathsf{P}} 0$. Next, note that

$$\begin{aligned}L_{1n}^* &\equiv s_{h_n}(X_0)^{-1}\langle(\tilde{\Gamma}_n^* - \tilde{\Gamma}_n)(\hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}), \Gamma_{h_n}^{-1}X_0\rangle^2 \\ &\le s_{h_n}(X_0)^{-1}\|\tilde{\Gamma}_n^* - \tilde{\Gamma}_n\|_\infty^2\|\Gamma_{h_n}^{-1}X_0\|^2\|\hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}\|^2.\end{aligned}$$

Due to the fact that $\mathsf{E}^*[\|\tilde{\Gamma}_n^* - \tilde{\Gamma}_n\|_\infty^2] = O_{\mathsf{P}}(n^{-1})$ as seen above, we have that

$$\mathsf{E}^*[s_{h_n}(X_0)^{-1}\|\tilde{\Gamma}_n^* - \tilde{\Gamma}_n\|_\infty^2\|\Gamma_{h_n}^{-1}X_0\rangle\|^2|X_0] = O_{\mathsf{P}}\left(n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right).$$

This implies that for each $\eta > 0$, $\mathsf{P}^*(L_{1n}^* > \eta) \xrightarrow{\mathsf{P}} 0$. Finally, note that

$$\begin{aligned}L_{2n}^* &\equiv s_{h_n}(X_0)^{-1}\langle\tilde{\Gamma}_n(\hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}), \Gamma_{h_n}^{-1}X_0\rangle^2 \\ &= s_{h_n}(X_0)^{-1}\|\tilde{\Gamma}_n\Gamma_{h_n}^{-1}X_0\|^2\|\hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}\|^2.\end{aligned}$$

As seen in the proof of Lemma 23, one can show that $s_{h_n}(X_0)^{-1}\|\tilde{\Gamma}_n\Gamma_{h_n}^{-1}X_0\|^2 = O_{\mathsf{P}}(1)$. This implies that for each $\eta > 0$, $\mathsf{P}^*(L_{2n}^* > \eta | X_0) \xrightarrow{\mathsf{P}} 0$. In summary, by the decomposition (3.69), we have the desired result. □

**Lemma 43.** *For each $\eta > 0$, as $n \to \infty$, we have the following.*

1. $\mathsf{P}^*\left(\left\|n^{-1}\sum_{i=1}^n (X_i^*)^{\otimes 2}\langle\hat{\beta}_{k_n} - \hat{\beta}_{g_n}, X_i^*\rangle^2\right\|_\infty > \eta\right) \xrightarrow{\mathsf{P}} 0.$
2. *If* $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2} = O(1)$, *then*
$$\mathsf{P}^*\left(s_{h_n}(X_0)^{-1}\left\langle\left(n^{-1}\sum_{i=1}^n (X_i^*)^{\otimes 2}\langle\hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}, X_i^*\rangle^2\right)\Gamma_{h_n}^{-1}X_0, (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1})X_0\right\rangle > \eta\middle| X_0\right) \xrightarrow{\mathsf{P}} 0.$$
3. $\mathsf{P}^*\left(s_{h_n}(X_0)^{-1}\left\langle\left(n^{-1}\sum_{i=1}^n (X_i^*)^{\otimes 2}\langle\hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}, X_i^*\rangle^2\right)\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\right\rangle > \eta\middle| X_0\right) \xrightarrow{\mathsf{P}} 0.$

*Proof.* Note that

$$\begin{aligned}&\left\|n^{-1}\sum_{i=1}^n (X_i^*)^{\otimes 2}\langle\hat{\beta}_{k_n} - \hat{\beta}_{g_n}, X_i^*\rangle^2\right\|_\infty \\ \le& n^{-1}\sum_{i=1}^n \|X_i^*\|^4\|\hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}\|^2 \\ \le& \left|n^{-1}\sum_{i=1}^n \|X_i^*\|^4 - \mathsf{E}[\|X\|^4]\right|\|\hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}\|^2 + \mathsf{E}[\|X\|^4]\|\hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}\|^2.\end{aligned}$$

To see the consistency of $\overline{\|X^*\|^4} \equiv n^{-1}\sum_{i=1}^n \|X_i^*\|^4$ for $\mathsf{E}[\|X\|^4]$, we follows the proof of Theorem 2.1 of [3] by using the techinical lemmas therein. Note that $d_4(X_i^*, X_i) \to 0$ almost surely by Lemma 8.4 of [3]. Define $\phi(x) = \|x\|^4$ for $x \in \mathbb{H}$ so that $d_1(\|X_i^*\|^4, \|X_i\|^4) \to 0$ almost surely by Lemma 8.5 of [3]. By Lemma 8.7 of [3], it then implies that

$$d_1(\overline{\|X^*\|^4}, \overline{\|X\|^4}) \le n^{-1}\sum_{i=1}^n d_1(\|X_i^*\|^4, \|X_i\|^4) = d_1(\|X_i^*\|^4, \|X_i^*\|^4) \to 0$$

almost surely, where $\overline{\|X\|^4} \equiv n^{-1}\sum_{i=1}^n \|X_i\|^4$. Since $\overline{\|X\|^4} \to \mathsf{E}[\|X\|^4]$ almost surely by the strong law of large numbers, it almost surely happens that for each $\eta > 0$,
$\mathsf{P}^*(|\overline{\|X^*\|^4} - \mathsf{E}[\|X\|^4]| > \eta) \to 0$ almost surely. Therefore, the first part follows.

For the second part, note that

$$\begin{aligned}
L_n^* \equiv & \left\langle \left( n^{-1}\sum_{i=1}^n (X_i^*)^{\otimes 2}\langle \hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}, X_i^*\rangle^2 \right) \Gamma_{h_n}^{-1} X_0, Q_n^* \right\rangle \\
=& n^{-1}\sum_{i=1}^n \langle X_i^*, \Gamma_{h_n}^{-1}X_0\rangle\langle X_i^*, Q_n^*\rangle\langle \hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}, X_i\rangle^2 \\
=& n^{-1}\sum_{i=1}^n \langle X_i^*, \Gamma_{h_n}^{-1}X_0\rangle\langle X_i^*, Q_n^*\rangle\langle (X_i^*)^{\otimes 2}(\hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}), \hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}\rangle \\
=& \left\langle n^{-1}\sum_{i=1}^n \langle X_i^*, \Gamma_{h_n}^{-1}X_0\rangle\langle X_i^*, Q_n^*\rangle (X_i^*)^{\otimes 2}(\hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}), \hat{\beta}^*_{k_n} - \hat{\beta}_{g_n} \right\rangle
\end{aligned}$$

where $Q_n^* \equiv ((\hat{\Gamma}^*_{h_n})^{-1} - \Gamma_{h_n}^{-1})X_0$. This implies that

$$|L_n^*| \leq \left( n^{-1}\sum_{i=1}^n \|X_i^*\|^3 |\langle X_i^*, \Gamma_{h_n}^{-1}X_0\rangle| \right) \|Q_n^*\| \|\hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}\|^2.$$

We note from Cauchy-Schwarz inequality that

$$\mathsf{E}^*[\|X_i^*\|^3 |\langle X_i^*, \Gamma_{h_n}^{-1}X_0\rangle| | X_0]^2 \leq \mathsf{E}^*[\|X_i^*\|^4]\mathsf{E}^*[\|X_i^*\|^2\langle X_i^*, \Gamma_{h_n}^{-1}X_0\rangle^2 | X_0].$$

Since $\mathsf{E}^*[\|X_i^*\|^4] = n^{-1}\sum_{i=1}^n \|X_i\|^4 \xrightarrow{\mathsf{P}} \mathsf{E}[\|X\|^4]$, we have $\mathsf{E}^*[\|X_i^*\|^4] = O_{\mathsf{P}}(1)$. Since the FPC scores $\xi_j$ are uncorrelated random variables with mean zero and variance $\gamma_j$, we have from the independence between $\mathcal{X}_n \equiv \{X_i\}_{i=1}^n$ and $X_0$ that

$$\begin{aligned}
&\mathsf{E}[\mathsf{E}^*[\|X_i^*\|^2\langle X_i^*, \Gamma_{h_n}^{-1}X_0\rangle^2 | X_0]] \\
=& \mathsf{E}\left[ n^{-1}\sum_{i=1}^n \|X_i\|^2\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2 \right] = \mathsf{E}[\|X_i\|^2\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2] \\
=& \sum_{j=1}^{h_n} \gamma_j^{-1}\mathsf{E}[\|X_i\|^2\langle X_i, \phi_j\rangle^2].
\end{aligned}$$

By Condition (A2) and Cauchy-Schwarz inequality, we see that

$$\mathsf{E}[\|X_i\|^2\langle X_i, \phi_j\rangle^2] \leq \mathsf{E}[\|X_i\|^4]^{1/2}\mathsf{E}[\langle X_i\phi_j\rangle^4]^{1/2} \leq C\gamma_j,$$

which implies that

$$\mathsf{E}^*\left[h_n^{1/2}s_{h_n}(X_0)^{-1}\left(n^{-1}\sum_{i=1}^{n}\|X_i^*\|^3|\langle X_i^*,\Gamma_{h_n}^{-1}X_0\rangle|\right)\Big|X_0\right]=O_{\mathsf{P}}(1).$$

Due to Lemma 31, we have

$$\mathsf{E}^*[h_n^{-1/2}\|Q_n^*\||X_0]=O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{2/3}\right)$$

where the corresponding remainder term is negligible by following the argument in Remark 15. Thus, the second part follows.

For the last part, note that

$$\begin{aligned}
L_n^* \equiv& \left\langle\left(n^{-1}\sum_{i=1}^{n}(X_i^*)^{\otimes 2}\langle\hat\beta_{k_n}^*-\hat\beta_{g_n},X_i^*\rangle^2\right)\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\right\rangle\\
=&n^{-1}\sum_{i=1}^{n}\langle X_i^*,\Gamma_{h_n}^{-1}X_0\rangle^2\langle\hat\beta_{k_n}^*-\hat\beta_{g_n},X_i^*\rangle^2\\
=&n^{-1}\sum_{i=1}^{n}\langle X_i^*,\Gamma_{h_n}^{-1}X_0\rangle^2\langle(X_i^*)^{\otimes 2}(\hat\beta_{k_n}^*-\hat\beta_{g_n}),\hat\beta_{k_n}^*-\hat\beta_{g_n}\rangle\\
=&\left\langle n^{-1}\sum_{i=1}^{n}\langle X_i^*,\Gamma_{h_n}^{-1}X_0\rangle^2(X_i^*)^{\otimes 2}(\hat\beta_{k_n}^*-\hat\beta_{g_n}),\hat\beta_{k_n}^*-\hat\beta_{g_n}\right\rangle.
\end{aligned}$$

This implies that

$$|L_n|\le\left(n^{-1}\sum_{i=1}^{n}\|X_i^*\|^2\langle X_i^*,\Gamma_{h_n}^{-1}X_0\rangle^2\right)\|\hat\beta_{k_n}^*-\hat\beta_{g_n}\|^2.$$

Since the FPC scores $\xi_j$ are uncorrelated random variables with mean zero and variance $\gamma_j$, we have from the independence between $\mathcal{X}_n\equiv\{X_i^*\}_{i=1}^n$ and $X_0$ that

$$\begin{aligned}
&\mathsf{E}[\mathsf{E}^*[\|X_i^*\|^2\langle X_i^*,\Gamma_{h_n}^{-1}X_0\rangle^2|X_0]]\\
=&\mathsf{E}\left[n^{-1}\sum_{i=1}^{n}\|X_i\|^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2\right]=\mathsf{E}[\|X_i\|^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2]\\
=&\sum_{j=1}^{h_n}\gamma_j^{-1}\mathsf{E}[\|X_i\|^2\langle X_i,\phi_j\rangle^2].
\end{aligned}$$

By Condition (A2) and Cauchy-Schwarz inequality, we see that

$$\mathsf{E}[\|X_i\|^2\langle X_i,\phi_j\rangle^2]\le\mathsf{E}[\|X_i\|^4]^{1/2}\mathsf{E}[\langle X_i,\phi_j\rangle^4]^{1/2}\le C\gamma_j,$$

which implies that

$$\mathsf{E}^*\left[s_{h_n}(X_0)^{-1}\left(n^{-1}\sum_{i=1}^n \|X_i^*\|^2\langle X_i^*,\Gamma_{h_n}^{-1}X_0\rangle^2\right)\Big|X_0\right]=O_{\mathsf{P}}(1).$$

Thus, we have the desired result. □

**Lemma 44.** *For each $\eta>0$, as $n\to\infty$, we have the following.*

1. $\mathsf{P}^*\left(\left\|n^{-1}\sum_{i=1}^n\{(X_i^*\varepsilon_{i,g_n}^*)\otimes X_i^*\}\langle\hat\beta_{k_n}^*-\hat\beta_{g_n},X_i^*\rangle\right\|_\infty>\eta\right)\xrightarrow{\mathsf{P}}0.$
2. *If $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}=O(1)$, then*

$$\mathsf{P}^*\left(\left\langle\left(n^{-1}\sum_{i=1}^n\{(X_i^*\varepsilon_{i,g_n}^*)\otimes X_i^*\}\langle\hat\beta_{k_n}^*-\hat\beta_{g_n},X_i\rangle\right)\Gamma_{h_n}^{-1}X_0,(\hat\Gamma_{h_n}^{-1}-\Gamma_{h_n}^{-1})X_0\right\rangle>\eta\Big|X_0\right)\xrightarrow{\mathsf{P}}0.$$

3. $\mathsf{P}^*\left(s_{h_n}(X_0)^{-1}\left|\left\langle\left(n^{-1}\sum_{i=1}^n\{(X_i^*\varepsilon_{i,g_n}^*)\otimes X_i\}\langle\hat\beta_{k_n}^*-\hat\beta_{g_n},X_i^*\rangle^2\right)\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\right\rangle\right|>\eta\Big|X_0\right)\xrightarrow{\mathsf{P}}$ 0.

*Proof.* We first observe that

$$\begin{aligned}&\left\|n^{-1}\sum_{i=1}^n\{(X_i^*\varepsilon_{i,g_n}^*)\otimes X_i^*\}\langle\hat\beta_{k_n}^*-\hat\beta_{g_n},X_i^*\rangle\right\|_\infty\\ \le&n^{-1}\sum_{i=1}^n\|X_i^*\varepsilon^*\|\|X_i^*\|^2\|\hat\beta_{k_n}^*-\hat\beta_{g_n}\|\\ \le&\left(n^{-1}\sum_{i=1}^n\|X_i^*\varepsilon^*\|^2\right)^{1/2}\left(n^{-1}\sum_{i=1}^n\|X_i^*\|^4\right)\|\hat\beta_{k_n}^*-\hat\beta_{g_n}\|\\ \le&C\left(n^{-1}\sum_{i=1}\|X_i^*Y_i^*\|^2+n^{-1}\sum_{i=1}^n\|X_i^*\|^2\|\hat\beta_{g_n}\|\right)^{1/2}\left(n^{-1}\sum_{i=1}^n\|X_i^*\|^4\right)\|\hat\beta_{k_n}^*-\hat\beta_{g_n}\|.\end{aligned}$$

As done in part 1 of Lemma 43, one can derive that for each $\eta>0$, $\mathsf{P}^*(|\overline{\|X^*Y^*\|^2}-\mathsf{E}[\|XY\|^2]|>\eta)\to0$, $\mathsf{P}^*(|\overline{\|X^*\|^2}-\mathsf{E}[\|X\|^2]|>\eta)\to0$, $\mathsf{P}^*(|\overline{\|X^*\|^4}-\mathsf{E}[\|X\|^4]|>\eta)\to0$ almost surely since $\mathsf{E}[\|XY\|^2]<\infty$, $\mathsf{E}[\|X\|^2]<\infty$, and $\mathsf{E}[\|X\|^4]<\infty$. This implies that

$$\mathsf{P}^*\left(\left\|n^{-1}\sum_{i=1}^n\{(X_i^*\varepsilon_{i,g_n}^*)\otimes X_i^*\}\langle\hat\beta_{k_n}^*-\hat\beta_{g_n},X_i^*\rangle\right\|_\infty>\eta\right)\xrightarrow{\mathsf{P}}0.$$

The last two parts follow from a similar argument to Lemmas 25 and 43. □

**Proposition 26.** *Suppose that as $n \to \infty$, $\|\hat{\beta}_{k_n} - \beta\| \xrightarrow{\mathsf{P}} 0$, $\|\hat{\beta}_{g_n} - \beta\| \xrightarrow{\mathsf{P}} 0$, and for each $\eta > 0$, $\mathsf{P}^*(\|\hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}\| > \eta) \xrightarrow{\mathsf{P}} 0$. As $n \to \infty$, if $n^{-1/2} h_n^{-1/2} \sum_{j=1}^{h_n} \delta_j^{-1/2} (j \log j)^{3/2} \to 0$, then the bootatrap scaling $\hat{s}^*_{h_n}(X_0)$ and the sample scaling $\hat{s}_{h_n}(X_0)$ are asymptotically equivalent in that, for any $\eta > 0$,*

$$\mathsf{P}^* \left( \left| \frac{\hat{s}^*_{h_n}(X_0)}{\hat{s}_{h_n}(X_0)} - 1 \right| > \eta \middle| X_0 \right) \xrightarrow{\mathsf{P}} 0.$$

*Proof.* Similarly to the inequality (3.46), we obtain the following decomposition from (3.67):

$$\begin{aligned}
&\left| \frac{\hat{s}^*_{h_n}(X_0)}{s_{h_n}(X_0)} - 1 \right| \\
\leq & s_{h_n}(X_0)^{-1} |\langle (\hat{\Lambda}^*_{n,k_n,g_n} - \Lambda) \{ (\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n} \} X_0, \{ (\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n} \} X_0 \rangle | \\
& + s_{h_n}(X_0)^{-1} |\langle \Lambda \{ (\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n} \} X_0, \{ (\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n} \} X_0 \rangle | \\
& + 2 s_{h_n}(X_0)^{-1} |\langle (\hat{\Lambda}^*_{n,k_n,g_n} - \Lambda) \Gamma^{-1}_{h_n} X_0, \{ (\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n} \} X_0 \rangle | \\
& + 2 s_{h_n}(X_0)^{-1} |\langle \Lambda \Gamma^{-1}_{h_n} X_0, \{ (\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n} \} X_0 \rangle | \\
& + s_{h_n}(X_0)^{-1} |\langle (\hat{\Lambda}^*_{n,k_n,g_n} - \Lambda) \Gamma^{-1}_{h_n} X_0, \Gamma^{-1}_{h_n} X_0 \rangle |.
\end{aligned} \tag{3.70}$$

Note from Lemma 41 and the first parts of Lemmas 42-44 that for each $\eta > 0$,

$$\mathsf{P}^*(\|\hat{\Lambda}^*_{n,k_n,g_n} - \Lambda\|_\infty > \eta) \xrightarrow{\mathsf{P}} 0.$$

The first two terms in (3.70) converges to zero by Lemmas 31 and 41, in the view of Remark 15, since for each $\eta > 0$,

$$\mathsf{P}^*(s_{h_n}(X_0)^{-1} |\langle (\hat{\Lambda}^*_{n,k_n,g_n} - \Lambda) \{ (\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n} \} X_0, \{ (\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n} \} X_0 \rangle | > \eta | X_0) \xrightarrow{\mathsf{P}} 0,$$
$$\mathsf{P}^*(s_{h_n}(X_0)^{-1} |\langle \Lambda \{ (\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n} \} X_0, \{ (\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n} \} X_0 \rangle | > \eta | X_0) \xrightarrow{\mathsf{P}} 0,$$

if $n^{-1/2} h_n^{-1/2} \sum_{j=1}^{h_n} \delta_j^{-1/2} (j \log j)^{3/2} \to 0$.

Define $\check{\Lambda}^*_n \equiv n^{-1} \sum_{i=1}^n (X^*_i Y^*_i - X^*_i \langle X^*_i, \beta \rangle)^{\otimes 2}$ so that $\mathsf{E}[\mathsf{E}^*[n \|\check{\Lambda}^*_n - \Lambda\|^2_{HS}]] = O(1)$ by Lemma 41. Note that

$$\begin{aligned}
& s_{h_n}(X_0)^{-1} |\langle (\check{\Lambda}^*_n - \Lambda) \Gamma^{-1}_{h_n} X_0, \{ (\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n} \} X_0 \rangle | \\
\leq & \{ h_n s_{h_n}(X_0)^{-1} \} (n^{1/2} \|\check{\Lambda}^*_n - \Lambda\|_\infty) (n^{-1/2} h_n^{-1/2} \|\Gamma^{-1}_{h_n} X_0\|) \{ h_n^{-1/2} \| (\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n} \} X_0 \| \}
\end{aligned}$$

Since $\mathsf{E}[\|\Gamma_{h_n}^{-1}X_0\|^2] = \sum_{j=1}^{h_n}\gamma_j^{-1}$ and $n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1} = O(1)$ holds by Condition (A5), if $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\gamma_j^{-1/2}(j\log j)^{3/2} \to 0$, by Lemma 31, we have

$$\mathsf{P}^*(s_{h_n}(X_0)^{-1}|\langle(\check{\Lambda}_n^* - \Lambda)\Gamma_{h_n}^{-1}X_0, \{(\hat{\Gamma}_{h_n}^*)^{-1} - \Gamma_{h_n}^{-1}\}X_0\rangle| > \eta|X_0) \xrightarrow{\mathsf{P}} 0.$$

One can show that

$$\mathsf{P}^*(s_{h_n}(X_0)^{-1}|\langle(\tilde{\Lambda}_{n,g_n}^* - \check{\Lambda}_n^*)\Gamma_{h_n}^{-1}X_0, \{(\hat{\Gamma}_{h_n}^*)^{-1} - \Gamma_{h_n}^{-1}\}X_0\rangle| > \eta|X_0) \xrightarrow{\mathsf{P}} 0$$

by using a similar argument to the proofs of the second parts of Lemmas 43-44 and interchanging $\hat{\beta}_{h_n}^* - \hat{\beta}_{g_n}$ into $\hat{\beta}_{g_n} - \beta$. This along with the second parts of Lemmas 42-44 implies that

$$\mathsf{P}^*(s_{h_n}(X_0)^{-1}|\langle(\hat{\Lambda}_{n,k_n,g_n}^* - \check{\Lambda}_n^*)\Gamma_{h_n}^{-1}X_0, \{(\hat{\Gamma}_{h_n}^*)^{-1} - \Gamma_{h_n}^{-1}\}X_0\rangle| > \eta|X_0) \xrightarrow{\mathsf{P}} 0.$$

Thus, the third term in (3.70) converges to zero.

The fourth term in (3.70) converges to zero by using the same argument to derive the converges of the fourth term in (3.46) to zero as seen in the proof of Proposition 20.

To deal with the last term in (3.70), note that

$$\begin{aligned}
&\mathsf{E}^*[s_{h_n}(X_0)^{-1}|\langle(\check{\Lambda}_n^* - \Lambda)\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle||X_0] \\
\le& s_{h_n}(X_0)^{-1}\mathsf{E}^*[\|\check{\Lambda}_n^* - \Lambda\|_\infty]\|\Gamma_{h_n}^{-1}X_0\|^2 \\
=& O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)
\end{aligned}$$

where $n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1} \to 0$ under Condition (A5). One can show that

$$\mathsf{P}^*(s_{h_n}(X_0)^{-1}|\langle(\tilde{\Lambda}_{n,g_n}^* - \check{\Lambda}_n^*)\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle| > \eta|X_0) \xrightarrow{\mathsf{P}} 0$$

by using a similar argument to the proofs of the second parts of Lemmas 43-44 and interchanging $\hat{\beta}_{h_n}^* - \hat{\beta}_{g_n}$ into $\hat{\beta}_{g_n} - \beta$. This along with the second parts of Lemmas 42-44 implies that

$$\mathsf{P}^*(s_{h_n}(X_0)^{-1}|\langle(\hat{\Lambda}_{n,k_n,g_n}^* - \check{\Lambda}_n^*)\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle| > \eta|X_0) \xrightarrow{\mathsf{P}} 0.$$

The above four arguments completes the proof along with the decomposition (3.70) and Proposition 20. □

*Remark* 16. The result in Proposition 26 still holds even when the truncation levels of the estimators $\hat{\beta}_{k_n}$ and $\hat{\beta}^*_{k_n}$ for computing residuals for $\hat{s}_{h_n}(X_0)$ and $\hat{s}^*_{h_n}(X_0)$ are not equal, as long as the estimators are consistent. For example, suppose that $\hat{\beta}_{k_n}$ and $\hat{\beta}^*_{k'_n}$ are respectively the estimators with distinct truncation levels $k_n$ and $k'_n$ used for constructing the scaling $\hat{s}_{h_n}(X_0)$ and $\hat{s}^*_{h_n}(X_0)$. Nevertheless, the bootstrap scaling $\hat{s}^*_{h_n}(X_0)$ is still consistent if both $\hat{\beta}_{k_n}$ and $\hat{\beta}^*_{k'_n}$ are consistent in the sense that as $n \to \infty$, $\|\hat{\beta}_{k_n} - \beta\| \xrightarrow{\mathsf{P}} 0$ and for each $\eta > 0$, $\mathsf{P}^*(\|\hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}\| > \eta) \xrightarrow{\mathsf{P}} 0$.

### 3.10.6 Failure of naive paired bootstrap

Note the following decomposition of the difference between the naive and our modified bootstrap estimator:

$$\hat{\beta}^*_{h_n,naive} - \hat{\beta}^*_{h_n} = (\hat{\Gamma}^*_{h_n})^{-1}\hat{U}_{n,g_n} = \{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}\hat{U}_{n,g_n} + \Gamma^{-1}_{h_n}\hat{U}_{n,g_n}.$$

The cross-covariance function $\hat{U}_{n,g_n} \equiv n^{-1}\sum_{i=1}^{n} X_i\hat{\varepsilon}_{i,g_n}$ between residuals and regressors can be further expanded as follows:

$$\begin{aligned}\hat{U}_{n,g_n} &= (I - \hat{\Pi}_{g_n})\hat{\Delta}_n = (I - \Pi_{g_n})\hat{\Delta}_n + (\Pi_{g_n} - \hat{\Pi}_{g_n})\hat{\Delta}_n \\ &= (I - \Pi_{g_n})U_n + (I - \Pi_{g_n})(\hat{\Gamma}_n - \Gamma)\beta + (I - \Pi_{g_n})\Gamma\beta \\ &\quad + (\Pi_{g_n} - \hat{\Pi}_{g_n})U_n + (\Pi_{g_n} - \hat{\Pi}_{g_n})(\hat{\Gamma}_n - \Gamma)\beta + (\Pi_{g_n} - \hat{\Pi}_{g_n})\Gamma\beta.\end{aligned} \tag{3.71}$$

The difference between the naive and our modified bootstrap statistics is then

$$T^*_{n,naive}(X_0) - T^*_n(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle(\hat{\Gamma}^*_{h_n})^{-1}\hat{U}_{n,g_n}, X_0\rangle = A^*_n + B_n + C_n,$$

where

$$A^*_n \equiv A^*_n(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\{(\hat{\Gamma}^*_{h_n})^{-1} - \Gamma^{-1}_{h_n}\}\hat{U}_{n,g_n}, X_0\rangle, \tag{3.72}$$

$$B_n \equiv B_n(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\Gamma^{-1}_{h_n}(I - \Pi_{g_n})\hat{\Delta}_n, X_0\rangle, \tag{3.73}$$

$$C_n \equiv C_n(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\Gamma^{-1}_{h_n}(\Pi_{g_n} - \hat{\Pi}_{g_n})\hat{\Delta}_n, X_0\rangle. \tag{3.74}$$

Here, the bootstrap statistics $T^*_{n,naive}(X_0)$ and $T^*_n(X_0)$ are written with the scaling $s_{h_n}(X_0)$ because it is enough to work with the scaling $s_{h_n}(X_0)$ thanks to Proposition 20. By Propositions 27-28, both terms $A^*_n$ and $C_n$ respectively from (3.72) and (3.74) converge to zero. Meanwhile, if $h_n \leq g_n$, we notice that $\Gamma^{-1}_{h_n}(I - \Pi_{g_n}) = 0$, which implies that the term in (3.73) is zero and the difference $T^*_{n,naive}(X_0) - T^*_n(X_0)$ hence converges to zero. In contrast, if $h_n > g_n$, the term in (3.73) then does not disappear and is expanded as

$$\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle \Gamma^{-1}_{h_n}(I - \Pi_{g_n})\hat{\Delta}_n, X_0\rangle = B_{1n} + B_{2n} + B_{3n},$$

where

$$B_{1n} \equiv B_{1n}(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle \Gamma^{-1}_{h_n}(I - \Pi_{g_n})U_n, X_0\rangle \tag{3.75}$$

$$B_{2n} \equiv B_{2n}(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle \Gamma^{-1}_{h_n}(I - \Pi_{g_n})(\hat{\Gamma}_n - \Gamma)\beta, X_0\rangle \tag{3.76}$$

$$B_{3n} \equiv B_{3n}(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle \Gamma^{-1}_{h_n}(I - \Pi_{g_n})\Gamma\beta, X_0\rangle. \tag{3.77}$$

We will show that the term $B_{3n}$ (3.77) converges to zero in Lemma 49 and that the term $B_{1n} + B_{2n}$ from (3.75) and (3.76) weakly converges to some normal random variable in Proposition 31.

#### 3.10.6.1 Convergence of $A^*_n$

The following lemma is a modification of Lemma 12, which is used for the convergence of $A^*_n$ from (3.72).

**Lemma 45.** *Under the same assumptions of Lemma 11, we have that*

$$\sum_{l\neq j}\frac{\gamma_l^2}{|\gamma_l - \gamma_j|} \leq C\gamma_j j \log j$$

*for sufficiently large $j \in \mathbb{N}$.*

*Proof.* We first decompose the sum into three terms

$$\sum_{l\neq j}\frac{\gamma_l^2}{|\gamma_l - \gamma_j|} = \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3,$$

where

$$\mathcal{T}_1 = \sum_{1\le l<j} \frac{\gamma_l^2}{|\gamma_l-\gamma_j|}, \quad \mathcal{T}_2 = \sum_{j<l\le 2j} \frac{\gamma_l^2}{|\gamma_l-\gamma_j|}, \quad \text{and} \quad \mathcal{T}_3 = \sum_{l>2j} \frac{\gamma_l^2}{|\gamma_l-\gamma_j|}.$$

For $l<j$, by the first part of Lemma 11,

$$\frac{\gamma_l^2}{|\gamma_l-\gamma_j|} = \gamma_j \frac{\gamma_l}{\gamma_j}\frac{\gamma_l}{\gamma_l-\gamma_j} \le \gamma_j \frac{j}{l}\frac{j}{j-l}$$

for sufficiently large $l$, and thus,

$$\begin{aligned}
\mathcal{T}_1 &= \sum_{1\le l<j} \frac{\gamma_l^2}{|\gamma_l-\gamma_j|} \le \gamma_j \sum_{1\le l<j} \frac{j}{l}\frac{j}{j-l} = \gamma_j j^2 \sum_{1\le l<j} \frac{1}{l(j-l)} \\
&= \gamma_j j \sum_{1\le l<j} \left(\frac{1}{l}+\frac{1}{j-l}\right) = 2\gamma_j j \sum_{1\le l<j} \frac{1}{l} \\
&\le C\gamma_j j \log j.
\end{aligned}$$

If $j<l\le 2j$, by the first part of Lemma 11, $j\gamma_j \ge l\gamma_l$ and

$$\frac{\gamma_l^2}{|\gamma_l-\gamma_j|} = \frac{\gamma_l^2}{\gamma_j-\gamma_l} = \gamma_j \frac{\gamma_l}{\gamma_j}\frac{\gamma_l}{\gamma_j}\frac{\gamma_j}{\gamma_j-\gamma_l} \le \gamma_j \left(\frac{j}{l}\right)^2 \frac{l}{l-j} = \frac{\gamma_j j^2}{l(l-j)}$$

for sufficiently large $l$. Thus, we have

$$\begin{aligned}
0 \le \mathcal{T}_2 &= \sum_{j<l\le 2j} \frac{\gamma_l^2}{|\gamma_l-\gamma_j|} \le \gamma_j j \sum_{j<l\le 2j} \frac{j}{l(l-j)} \\
&\le \gamma_j j \sum_{j<l\le 2j} \left(-\frac{1}{l}+\frac{1}{l-j}\right) = \gamma_j j \left(\sum_{l=1}^{j} \frac{1}{l} - \sum_{j<l\le 2j} \frac{1}{l}\right) \\
&\le C\gamma_j j \log j.
\end{aligned}$$

For $l>2j$, since $\gamma_l \le \gamma_{2j}$,

$$\mathcal{T}_3 = \sum_{l>2j} \frac{\gamma_l^2}{|\gamma_l-\gamma_j|} \le \frac{\sum_{l>2j}\gamma_l^2}{\gamma_j-\gamma_{2j}} \le \gamma_{2j}\frac{\sum_{l>2j}\gamma_l}{\gamma_j-\gamma_{2j}} \le \gamma_j \frac{\sum_{l>2j}\gamma_l}{\gamma_j-\gamma_{2j}}$$

Again by the first part of Lemma 11,

$$\frac{1}{\gamma_j-\gamma_{2j}} \le \frac{2j}{2j-j}\frac{1}{\gamma_j} = \frac{2}{\gamma_j}$$

and by the second part of Lemma 11, $\sum_{l>2j} \gamma_j \leq (2j+1)\gamma_{2j}$. This implies that

$$\mathcal{T}_3 \leq \gamma_j 2(2j+1)\frac{\gamma_{2j}}{\gamma_j}.$$

Finally, again by the first part of Lemma 11, we have $\gamma_j \geq 2\gamma_{2j}$, which implies that

$$\mathcal{T}_3 \leq \gamma_j(2j+1) \leq C\gamma_j j \leq C\gamma_j j \log j.$$

□

The term $A_n^*$ in (3.72) converges to zero as follows.

**Proposition 27.** *Suppose that Conditions (A1)-(A6) and (A8) hold. As $n \to \infty$, if*

$$h_n^{-1} + g_n^{-1} + n^{-1/2}h_n^2(\log h_n)^{3/2}g_n^2(\log g_n) + n^{-1/2}h_n^{7/2}(\log h_n)^3 \to 0,$$

*then, for each $\eta > 0$, we have $\mathsf{P}^*(|A_n^*| > \eta | X_0) \xrightarrow{\mathsf{P}} 0$.*

*Proof.* Based on the decomposition of $\hat{U}_{n,g_n} \equiv n^{-1}\sum_{i=1}^n X_i\hat{\varepsilon}_{i,g_n}$ in (3.71), $A_n^*$ can be further decomposed as $A_n^* = \sum_{l=1}^6 A_{ln}^*$, where

$$A_{1n}^* \equiv A_{1n}^*(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\{(\hat{\Gamma}_{h_n}^*)^{-1} - \Gamma_{h_n}^{-1}\}(I - \Pi_{g_n})U_n, X_0\rangle, \tag{3.78}$$

$$A_{2n}^* \equiv A_{2n}^*(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\{(\hat{\Gamma}_{h_n}^*)^{-1} - \Gamma_{h_n}^{-1}\}(I - \Pi_{g_n})(\hat{\Gamma}_n - \Gamma)\beta, X_0\rangle, \tag{3.79}$$

$$A_{3n}^* \equiv A_{3n}^*(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\{(\hat{\Gamma}_{h_n}^*)^{-1} - \Gamma_{h_n}^{-1}\}(I - \Pi_{g_n})\Gamma\beta\rangle, \tag{3.80}$$

$$A_{4n}^* \equiv A_{4n}^*(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\{(\hat{\Gamma}_{h_n}^*)^{-1} - \Gamma_{h_n}^{-1}\}(\hat{\Pi}_{g_n} - \Pi_{g_n})U_n, X_0\rangle, \tag{3.81}$$

$$A_{5n}^* \equiv A_{5n}^*(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\{(\hat{\Gamma}_{h_n}^*)^{-1} - \Gamma_{h_n}^{-1}\}(\hat{\Gamma}_n - \Gamma)(\hat{\Pi}_{g_n} - \Pi_{g_n})\beta, X_0\rangle, \tag{3.82}$$

$$A_{6n}^* \equiv A_{6n}^*(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\{(\hat{\Gamma}_{h_n}^*)^{-1} - \Gamma_{h_n}^{-1}\}\Gamma(\hat{\Pi}_{g_n} - \Pi_{g_n})\beta, X_0\rangle. \tag{3.83}$$

Following the spirit of Lemma 17 and Remark 15, we ignore the remainder terms related to either $\mathcal{E}_j^c$, $\mathcal{A}_{h_n}^c$, $(\mathcal{E}_j^*)^c$, or $(\mathcal{A}_{h_n}^*)^c$.

The term $A_{1n}^*$ in (3.78) converges to zero as follows. One can show that

$$\begin{aligned}\mathsf{E}[\|(I - \Pi_{h_n})U_n\|^2] &\leq C\left(n^{-1}\sum_{j>g_n}\langle\Lambda\phi_j, \phi_j\rangle + n^{-2}\sum_{j>g_n}\gamma_j\right) \\ &= o(n^{-1})\end{aligned}$$

by a similar argument to the one in the proof of Lemma 22. This implies that

$$\begin{aligned}
&\mathsf{E}^*\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle\{(\hat{\Gamma}^*_{h_n})^{-1}-\Gamma^{-1}_{h_n}\}(I-\Pi_{g_n})U_n, X_0\rangle|\right]\\
\le&\sqrt{\frac{n}{s_{h_n}(X_0)}}\|(I-\Pi_{g_n})U_n\|\mathsf{E}^*[\|\{(\hat{\Gamma}^*_{h_n})^{-1}-\Gamma^{-1}_{h_n}\}X_0\|]\\
=&o_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)
\end{aligned}$$

from Lemma 31.

The second term $A^*_{2n}$ in (3.79) is bounded above as

$$\begin{aligned}
&\mathsf{E}^*\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle\{(\hat{\Gamma}^*_{h_n})^{-1}-\Gamma^{-1}_{h_n}\}(\hat{\Gamma}_n-\Gamma)(\Pi_{h_n}-I)\beta, X_0\rangle|\right]\\
\le&\sqrt{\frac{n}{s_{h_n}(X_0)}}\|I-\Pi_{g_n}\|_\infty\|\hat{\Gamma}_n-\Gamma\|_\infty\|\beta\|\mathsf{E}^*[\|\{(\hat{\Gamma}^*_{h_n})^{-1}-\Gamma^{-1}_{h_n}\}X_0\|]\\
=&O_{\mathsf{P}}(n^{1/2}h_n^{-1/2})O_{\mathsf{P}}(n^{-1/2})O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)\\
=&O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right).
\end{aligned}$$

Showing the convergence of the term $A^*_{3n}$ in (3.80) needs more efforts. We will use a similar techniques to prove Proposition 23. By applying the second resolvent identity (Lemma 14) twice, we have

$$\begin{aligned}
&(zI-\hat{\Gamma}^*_n)^{-1}-(zI-\Gamma)^{-1}\\
=&\{(zI-\Gamma)^{-1}+(zI-\hat{\Gamma}^*_n)^{-1}(\hat{\Gamma}^*_n-\Gamma)(zI-\Gamma)^{-1}\}(\hat{\Gamma}^*_n-\Gamma)(zI-\Gamma)^{-1}\\
=&(zI-\Gamma)^{-1}(\hat{\Gamma}^*_n-\Gamma)(zI-\Gamma)^{-1}+(zI-\hat{\Gamma}^*_n)^{-1}(\hat{\Gamma}^*_n-\Gamma)(zI-\Gamma)^{-1}(\hat{\Gamma}^*_n-\Gamma)(zI-\Gamma)^{-1}\\
=&(zI-\Gamma)^{-1}(\hat{\Gamma}^*_n-\Gamma)(zI-\Gamma)^{-1}+(zI-\Gamma)^{-1}(\hat{\Gamma}^*_n-\Gamma)(zI-\Gamma)^{-1}(\hat{\Gamma}^*_n-\Gamma)(zI-\hat{\Gamma}^*_n)^{-1}
\end{aligned}$$

since all quantities are symmetric. This implies that

$$\begin{aligned}
(\hat{\Gamma}^*_{h_n})^{-1}-\Gamma^{-1}_{h_n}&=\frac{1}{2\pi\iota}\int_{\mathcal{C}_{h_n}}z^{-1}\left\{(zI-\hat{\Gamma}^*_n)^{-1}-(zI-\Gamma)^{-1}\right\}dz+r_{1n}\mathbb{I}_{(\mathcal{A}^*_{h_n})^c}\\
&=\mathcal{S}^*_n+\mathcal{R}^*_n+r^*_{2n}\mathbb{I}_{(\mathcal{A}^*_{h_n})^c}
\end{aligned}$$

where

$$\check{\mathcal{S}}_n^* = \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}(zI-\Gamma)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\Gamma)^{-1}dz,$$
$$\check{\mathcal{R}}_n^* = \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}(zI-\Gamma)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\Gamma)^{-1}(\hat{\Gamma}_n^*-\Gamma)(zI-\hat{\Gamma}_n^*)^{-1}dz,$$

and $r_{2n}^* = (\hat{\Gamma}_{h_n}^*)^{-1} - \frac{1}{2\pi\iota}\int_{\mathcal{C}_{h_n}} z^{-1}(zI-\hat{\Gamma}_n^*)^{-1}dz$ (cf. Lemma 29).

For the first term $\check{\mathcal{S}}_n^*$, we use a similar argument to that in Lemma 35. Then, one can derive that

$$\mathsf{E}[\mathsf{E}^*[\langle\check{\mathcal{S}}_n^*(I-\Pi_{g_n})\Gamma\beta, X_0\rangle^2]] = \sum_{l=1}^{\infty}\gamma_l\mathsf{E}[\mathsf{E}^*[\langle\check{\mathcal{S}}_n^*(I-\Pi_{g_n})\Gamma\beta,\phi_l\rangle^2]]$$

and $\langle\check{\mathcal{S}}_n^*(I-\Pi_{g_n})\Gamma\beta,\phi_l\rangle = \sum_{l'>g_n}\gamma_{l'}\beta_{l'}\langle\check{\mathcal{S}}_n^*\phi_l,\phi_l'\rangle$ with

$$\langle\check{\mathcal{S}}_n^*\phi_l,\phi_l'\rangle = \begin{cases} \frac{\langle(\hat{\Gamma}_n^*-\Gamma)\phi_l,\phi_{l'}\rangle}{\gamma_l(\gamma_l-\gamma_{l'})} & \text{if } l\le h_n\le g_n<l' \\ \frac{\langle(\hat{\Gamma}_n^*-\Gamma)\phi_l,\phi_{l'}\rangle}{\gamma_{l'}(\gamma_{l'}-\gamma_l)} & \text{if } g_n<l'\le h_n<l \\ 0 & \text{otherwise.}\end{cases}$$

This implies that

$$\begin{aligned}
&\mathsf{E}[\mathsf{E}^*[\langle\check{\mathcal{S}}_n^*(I-\Pi_{g_n})\Gamma\beta, X_0\rangle^2]] \\
&\le \begin{cases} Cn^{-1}\sum_{l=1}^{h_n}\gamma_l\left(\sum_{l'>g_n}|\beta_{l'}|\gamma_{l'}\frac{\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_l(\gamma_l-\gamma_{l'})}\right)^2 & \text{if } h_n\le g_n \\ Cn^{-1}\sum_{l>h_n}\gamma_l\left(\sum_{l'=1}^{g_n}|\beta_{l'}|\gamma_{l'}\frac{\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_{l'}(\gamma_{l'}-\gamma_l)}\right)^2 & \text{if } h_n> g_n\end{cases} \\
&\le \begin{cases} Cn^{-1}\sum_{l=1}^{h_n}\gamma_l\left(\sum_{l'>g_n}|\beta_{l'}|\frac{\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_l-\gamma_{l'}}\right)^2 & \text{if } h_n\le g_n \\ Cn^{-1}\sum_{l>h_n}\gamma_l\left(\sum_{l'=1}^{g_n}|\beta_{l'}|\frac{\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_{l'}-\gamma_l}\right)^2 & \text{if } h_n> g_n\end{cases}
\end{aligned}$$

since $\gamma_{l'}\le\gamma_l$ if $l\le h_n\le g_n<l'$. Hence, we conclude that

$$\mathsf{E}^*\left[\frac{n}{s_{h_n}(X_0)}\langle\check{\mathcal{S}}_n^*(I-\Pi_{g_n})\Gamma\beta, X_0\rangle^2\right] = o_{\mathsf{P}}(1).$$

For the second term $\check{\mathcal{R}}_n^*$, as in the proof of Lemma 36, we first see that

$$\check{\mathcal{R}}_n^* = \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}(zI-\Gamma)^{-1/2}G_n^*(z)^2K_n^*(z)(zI-\Gamma)^{-1/2}dz$$

and $|\langle \check{\mathcal{R}}_n^*(I-\Pi_{g_n})\Gamma\beta, X_0\rangle| \le C\sum_{j=1}^{h_n} A_j$ where

$$A_j = \int_{\mathcal{B}_j} |z|^{-1}\|K_n^*(z)\|_\infty \|G_n^*(z)\|_\infty^2 \|(zI-\Gamma)^{-1/2}(I-\Pi_{g_n})\Gamma\beta\|\|(zI-\Gamma)^{-1/2}X_0\|dz.$$

By Lemmas 11 and 45, for $z \in \mathcal{B}_j$, we have

$$\begin{aligned}
&\|(zI-\Gamma)^{-1/2}(I-\Pi_{g_n})\Gamma\beta\|^2 \\
=&\sum_{l>g_n} \frac{\gamma_l^2\beta_l^2}{|z-\gamma_l|} \le \sum_{l=1}^{\infty} \frac{\gamma_l^2\beta_l^2}{|z-\gamma_l|} \\
\le& C\left(\sum_{l\neq j} \frac{\gamma_l^2}{|z-\gamma_l|} + \frac{\gamma_j^2}{|z-\gamma_j|}\right) \le C\left(\sum_{l\neq j} \frac{\gamma_l^2}{|\gamma_j-\gamma_l|} + \frac{\gamma_j^2}{\delta_j}\right) \\
\le& C(\gamma_j j\log j + \gamma_j(j+1)) \le C\gamma_j j \log j.
\end{aligned}$$

This implies that

$$\begin{aligned}
\mathsf{E}[\mathsf{E}^*[A_j^*\mathbb{I}_{\mathcal{E}_j^*}]] &\le C\delta_j n^{-1}(j\log j)^2(\gamma_j j\log j)^{1/2}(j\log j)^{1/2} \\
&\le Cn^{-1}(j\log j)^{3/2}.
\end{aligned}$$

Thus, if $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}(j\log j)^{3/2} \le Cn^{-1/2}h_n^2(\log h_n)^{3/2} \to 0$, for each $\eta > 0$, we have

$$\mathsf{P}^*\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle \check{\mathcal{R}}_n^*(I-\Pi_{g_n})\Gamma\beta, X_0\rangle| > \eta\right) \xrightarrow{\mathsf{P}} 0.$$

The fourth term $A_{4n}^*$ is bounded as follows. One can show that $\|(\hat{\Pi}_{g_n}-\Pi_{g_n})U_n\| = O_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{g_n} j\log j\right)$ by a similar argument in the proof of Lemma 20. This implies that

$$\begin{aligned}
&\mathsf{E}^*\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle\{(\hat{\Gamma}_{h_n}^*)^{-1}-\Gamma_{h_n}^{-1}\}(\hat{\Pi}_{g_n}-\Pi_{g_n})U_n, X_0\rangle|\right] \\
\le&\sqrt{\frac{n}{s_{h_n}(X_0)}}\|(\hat{\Pi}_{g_n}-\Pi_{g_n})U_n\|\mathsf{E}^*[\|\{(\hat{\Gamma}_{h_n}^*)^{-1}-\Gamma_{h_n}^{-1}\}X_0\|] \\
=&O_{\mathsf{P}}\left(\left(n^{-1/2}\sum_{j=1}^{g_n} j\log j\right)\left\{n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right\}\right).
\end{aligned}$$

From Lemmas 20 and 31, the fifth term $A^*_{5n}$ is bounded as

$$
\begin{aligned}
&\mathsf{E}^*\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle\{(\hat{\Gamma}^*_{h_n})^{-1}-\Gamma^{-1}_{h_n}\}(\hat{\Gamma}_n-\Gamma)(\hat{\Pi}_{g_n}-\Pi_{g_n})\beta,X_0\rangle|\right]\\
\leq&\sqrt{\frac{n}{s_{h_n}(X_0)}}\|\hat{\Gamma}_n-\Gamma\|_\infty\|\hat{\Pi}_{g_n}-\Pi_{g_n}\|_\infty\mathsf{E}^*[\|\{(\hat{\Gamma}^*_{h_n})^{-1}-\Gamma^{-1}_{h_n}\}X_0\|]\\
=&O_{\mathsf{P}}(n^{1/2}h_n^{-1/2})O_{\mathsf{P}}(n^{-1/2})O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{g_n}j\log j\right)O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)\\
=&O_{\mathsf{P}}\left(\left(n^{-1/2}\sum_{j=1}^{g_n}j\log j\right)\left\{n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right\}\right).
\end{aligned}
$$

The last term $A^*_{6n}$ is bounded as follows. By Lemmas 11 and 45, it follows that

$$
\begin{aligned}
&\|\Gamma(zI-\Gamma)^{-1/2}\|^2_{HS}=\sum_{l=1}^{\infty}\frac{\gamma_l^2}{|z-\gamma_l|^2}\\
\leq&C\left(\sum_{l\neq j}\frac{\gamma_l^2}{|z-\gamma_l|}+\frac{\gamma_j^2}{|z-\gamma_j|}\right)\leq C\left(\sum_{l\neq j}\frac{\gamma_l^2}{|\gamma_j-\gamma_l|}+\frac{\gamma_j^2}{\delta_j}\right)\\
\leq&C(\gamma_j j\log j+\gamma_j(j+1))\leq C\gamma_j j\log j.
\end{aligned}
$$

By using the same argument as the one in Lemma 31, we then derive

$$
\mathsf{E}^*[\|\Gamma\{(\hat{\Gamma}^*_{h_n})^{-1}-\Gamma^{-1}_{h_n}\}X_0\|]=O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n}(j\log j)^{3/2}\right).
$$

Finally, by Lemma 20, we have

$$
\begin{aligned}
&\mathsf{E}^*\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle\{(\hat{\Gamma}^*_{h_n})^{-1}-\Gamma^{-1}_{h_n}\}\Gamma(\hat{\Pi}_{g_n}-\Pi_{g_n})\beta,X_0\rangle|\right]\\
\leq&\sqrt{\frac{n}{s_{h_n}(X_0)}}\|\hat{\Pi}_{g_n}-\Pi_{g_n}\|_\infty\mathsf{E}^*[\|\Gamma\{(\hat{\Gamma}^*_{h_n})^{-1}-\Gamma^{-1}_{h_n}\}X_0\|]\\
=&O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\left(\sum_{j=1}^{g_n}j\log j\right)\sum_{j=1}^{h_n}(j\log j)^{3/2}\right),
\end{aligned}
$$

where the last upper bound is bounded by $n^{-1/2}h_n^2(\log h_n)^{3/2}g_n^2(\log g_n)$. □

#### 3.10.6.2 Convergence of $C_n$

The term $C_n$ in (3.74) converges to zero as follows.

**Proposition 28.** *Suppose that Conditions (A1)-(A6) hold. As $n \to \infty$, if*

$$h_n^{-1} + g_n^{-1} + n^{-1/2} h_n^{-1/2} g_n^5 (\log g_n)^6 \to 0,$$

*then, for each $\eta > 0$, we have $\mathsf{P}(|C_n| > \eta | X_0) \xrightarrow{\mathsf{P}} 0$.*

*Proof.* The term $C_n$ in (3.74) can be further expanded as $C_n = C_{1n} + C_{2n} + C_{3n}$, where

$$C_{1n} \equiv C_{1n}(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}} \langle \Gamma_{h_n}^{-1} (\Pi_{g_n} - \hat{\Pi}_{g_n}) U_n, X_0 \rangle, \tag{3.84}$$

$$C_{2n} \equiv C_{2n}(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}} \langle \Gamma_{h_n}^{-1} (\Pi_{g_n} - \hat{\Pi}_{g_n}) (\hat{\Gamma}_n - \Gamma) \beta, X_0 \rangle, \tag{3.85}$$

$$C_{3n} \equiv C_{3n}(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}} \langle \Gamma_{h_n}^{-1} (\Pi_{g_n} - \hat{\Pi}_{g_n}) \Gamma \beta, X_0 \rangle. \tag{3.86}$$

Following the spirit of Lemma 17 and Remark 15, we ignore the remainder terms related to either $\mathcal{E}_j^c$, $\mathcal{A}_{h_n}^c$, $(\mathcal{E}_j^*)^c$, or $(\mathcal{A}_{h_n}^*)^c$.

The term $C_{1n}$ in (3.84) converges to zero as follows. One can show that $\|(\Pi_{g_n} - \hat{\Pi}_{g_n}) U_n\| = O_{\mathsf{P}}\left(n^{-1} \sum_{j=1}^{g_n} j \log j\right)$ by a similar argument in the proof of Lemma 20. This implies that

$$\begin{aligned}
&\mathsf{E}\left[\sqrt{\frac{n}{s_{h_n}(X_0)}} |\langle \Gamma_{h_n}^{-1} (\Pi_{g_n} - \hat{\Pi}_{g_n}) U_n, X_0 \rangle| \middle| X_0\right] \\
\leq& \sqrt{\frac{n}{s_{h_n}(X_0)}} \mathsf{E}[\|(\Pi_{g_n} - \hat{\Pi}_{g_n}) U_n\|] \|\Gamma_{h_n}^{-1} X_0\| \\
=& O_{\mathsf{P}}\left(\left(n^{-1/2} \sum_{j=1}^{g_n} j \log j\right) \left(h_n^{-1} \sum_{j=1}^{h_n} \gamma_j^{-1}\right)^{1/2}\right).
\end{aligned}$$

By Cauchy-Schwarz inequality, we have $\sum_{j=1}^{h_n} \gamma_j^{-1} \leq h_n^{1/2} \left(\sum_{j=1}^{h_n} \gamma_j^{-2}\right)^{1/2}$, which implies that $n^{-1/4} h_n^{-1/4} \left(\sum_{j=1}^{h_n} \gamma_j^{-1}\right)^{1/2} \leq \left(n^{-1} \sum_{j=1}^{h_n} \gamma_j^{-2}\right)^{1/4}$. Thus, by Condition (A5), we have

$$\begin{aligned}
&\mathsf{E}\left[\sqrt{\frac{n}{s_{h_n}(X_0)}} |\langle \Gamma_{h_n}^{-1} (\Pi_{g_n} - \hat{\Pi}_{g_n}) U_n, X_0 \rangle| \middle| X_0\right] \\
=& o_{\mathsf{P}}\left(\left(n^{-1/4} h_n^{-1/4} \sum_{j=1}^{g_n} j \log j\right)\right),
\end{aligned}$$

where the upper bound is dominated by $n^{-1/2} h_n^{-1/2} g_n^4 (\log g_n)^2$.

The term $C_{2n}$ in (3.85) converges to zero as follows. By using the same argument as the one in Lemma 20, we can show that $\mathsf{E}[\|(\Pi_{g_n}-\hat{\Pi}_{g_n})(\hat{\Gamma}_n-\Gamma)\beta\||X_0] = O_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{g_n} j\log j\right)$. We then have

$$
\begin{aligned}
&\mathsf{E}\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle\Gamma_{h_n}^{-1}(\Pi_{g_n}-\hat{\Pi}_{g_n})(\hat{\Gamma}_n-\Gamma)\beta, X_0\rangle|\Big|X_0\right]\\
\leq&\sqrt{\frac{n}{s_{h_n}(X_0)}}\mathsf{E}[\|(\Pi_{g_n}-\hat{\Pi}_{g_n})(\hat{\Gamma}_n-\Gamma)\beta\|]\|\Gamma_{h_n}^{-1}X_0\|\\
=&O_{\mathsf{P}}(n^{1/2}h_n^{-1/2})O_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{g_n} j\log j\right)O_{\mathsf{P}}\left(\left(\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2}\right)\\
=&O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{g_n} j\log j\left(h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2}\right).
\end{aligned}
$$

The last upper bound is the same upper bound for $C_{1n}$ above, which is again dominated by $n^{-1/2}h_n^{-1/2}g_n^4(\log g_n)^2$.

For the convergence of $C_{3n}$ from (3.86), we will use a similar technique to prove Proposition 17. By applying the second resolvent identity (Lemma 14) twice, we have

$$
\begin{aligned}
&(zI-\hat{\Gamma}_n)^{-1}-(zI-\Gamma)^{-1}\\
=&\{(zI-\Gamma)^{-1}+(zI-\hat{\Gamma}_n)^{-1}(\hat{\Gamma}_n-\Gamma)(zI-\Gamma)^{-1}\}(\hat{\Gamma}_n-\Gamma)(zI-\Gamma)^{-1}\\
=&(zI-\Gamma)^{-1}(\hat{\Gamma}_n-\Gamma)(zI-\Gamma)^{-1}+(zI-\hat{\Gamma}_n)^{-1}(\hat{\Gamma}_n-\Gamma)(zI-\Gamma)^{-1}(\hat{\Gamma}_n-\Gamma)(zI-\Gamma)^{-1}\\
=&(zI-\Gamma)^{-1}(\hat{\Gamma}_n-\Gamma)(zI-\Gamma)^{-1}+(zI-\Gamma)^{-1}(\hat{\Gamma}_n-\Gamma)(zI-\Gamma)^{-1}(\hat{\Gamma}_n-\Gamma)(zI-\hat{\Gamma}_n)^{-1}
\end{aligned}
$$

since all quantities are symmetric. This implies that

$$
\begin{aligned}
\hat{\Pi}_{g_n}-\Pi_{g_n} &= \frac{1}{2\pi\iota}\int_{\mathcal{C}_{g_n}}\left\{(zI-\hat{\Gamma}_n)^{-1}-(zI-\Gamma)^{-1}\right\}dz + r_{1n}\mathbb{I}_{\mathcal{A}_{h_n}^c}\\
&= \mathcal{S}_n+\mathcal{R}_n+r_{1n}\mathbb{I}_{\mathcal{A}_{g_n}^c}
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{S}_n &= \frac{1}{2\pi\iota}\sum_{j=1}^{g_n}\int_{\mathcal{B}_j}(zI-\Gamma)^{-1}(\hat{\Gamma}_n-\Gamma)(zI-\Gamma)^{-1}dz,\\
\mathcal{R}_n &= \frac{1}{2\pi\iota}\sum_{j=1}^{g_n}\int_{\mathcal{B}_j}(zI-\Gamma)^{-1}(\hat{\Gamma}_n-\Gamma)(zI-\Gamma)^{-1}(\hat{\Gamma}_n-\Gamma)(zI-\hat{\Gamma}_n)^{-1}dz,
\end{aligned}
$$

and $r_{1n} = \hat{\Pi}_{g_n} - \frac{1}{2\pi\iota}\int_{\mathcal{C}_{g_n}}(zI - \hat{\Gamma}_n)^{-1}dz$ (cf. Lemma 17). Proposition 17 For the first term $\mathcal{S}_n$, we use a similar argument to that in the proof of Proposition 2 in [12]. One can derive that

$$\mathsf{E}[\langle \mathcal{S}_n\Gamma\beta, \Gamma_{h_n}^{-1}X_0\rangle^2] = \mathsf{E}\left[\left(\sum_{l=1}^{h_n}\gamma_l^{-1}\langle X_0,\phi_l\rangle\langle \mathcal{S}_n\Gamma\beta,\phi_l\rangle\right)^2\right] = \sum_{l=1}^{h_n}\gamma_l^{-1}\mathsf{E}[\langle \mathcal{S}_n\Gamma\beta,\phi_l\rangle^2]$$

and

$$\begin{aligned}\mathsf{E}[\langle \mathcal{S}_n\Gamma\beta,\phi_l\rangle^2] &= \mathsf{E}\left[\left(\sum_{l'=1}^{\infty}\beta_{l'}\langle \mathcal{S}_n\Gamma\phi_{l'},\phi_l\rangle\right)^2\right] = \mathsf{E}\left[\left(\sum_{l'=1}^{\infty}\beta_{l'}\gamma_{l'}\langle \mathcal{S}_n\phi_{l'},\phi_l\rangle\right)^2\right]\\ &= \mathsf{E}\left[\left(\sum_{l'>h_n}^{\infty}\beta_{l'}\gamma_{l'}\frac{\langle(\hat{\Gamma}_n-\Gamma)\phi_{l'},\phi_l}{\gamma_l-\gamma_{l'}}\right)^2\right]\\ &\le C\left(\sum_{l'>h_n}|\beta_{l'}|\gamma_{l'}\frac{\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_l-\gamma_{l'}}\right)^2\end{aligned}$$

This implies that

$$\mathsf{E}[\langle \mathcal{S}_n\Gamma\beta, \Gamma_{h_n}^{-1}X_0\rangle^2] \le C\sum_{l=1}^{h_n}\gamma_l^{-1}\left(\sum_{l'>h_n}|\beta_{l'}|\gamma_{l'}\frac{\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_l-\gamma_{l'}}\right)^2 \le C\sum_{l=1}^{h_n}\gamma_l\left(\sum_{l'>h_n}|\beta_{l'}|\frac{\gamma_l^{1/2}\gamma_{l'}^{1/2}}{\gamma_l-\gamma_{l'}}\right)^2$$

since $\gamma_{l'} \le \gamma_l$ if $l \le h_n < l'$. Hence, we conclude that

$$\mathsf{E}\left[\frac{n}{s_{h_n}(X_0)}\langle \mathcal{S}_n\Gamma\beta, \Gamma_{h_n}^{-1}X_0\rangle^2\Big|X_0\right] = o_{\mathsf{P}}(1).$$

For the second term, as in the proof of Proposition 2 in [12], we first see that

$$\mathcal{R}_n = \frac{1}{2\pi\iota}\sum_{j=1}^{g_n}\int_{\mathcal{B}_j}(zI-\Gamma)^{-1/2}G_n(z)^2K_n(z)(zI-\Gamma)^{-1/2}dz$$

and $|\langle \mathcal{R}_n\Gamma\beta, \Gamma_{g_n}^{-1}X_0\rangle| \le C\sum_{j=1}^{h_n}A_j$ where

$$A_j = \int_{\mathcal{B}_j}\|G_n(z)\|_\infty^2\|K_n(z)\|_\infty\|(zI-\Gamma)^{-1/2}\Gamma\beta\|\|(zI-\Gamma)^{-1/2}\|\|\Gamma_{h_n}^{-1}X_0\|dz.$$

By Lemmas 11 and 45, for $z \in \mathcal{B}_j$, we have

$$\begin{aligned}\|(zI-\Gamma)^{-1/2}\Gamma\beta\|^2 &= \sum_{l=1}^{\infty}\frac{\gamma_l^2\beta_l^2}{|z-\gamma_l|} \le C\left(\sum_{l\ne j}\frac{\gamma_l^2}{|z-\gamma_l|}+\frac{\gamma_j^2}{|z-\gamma_j|}\right) \le C\left(\sum_{l\ne j}\frac{\gamma_l^2}{|\gamma_j-\gamma_l|}+\frac{\gamma_j^2}{\delta_j}\right)\\ &\le C(\gamma_j j\log j + \gamma_j(j+1)) \le C\gamma_j j\log j.\end{aligned}$$

This implies that

$$\mathsf{E}[A_j \mathbb{I}_{\mathcal{E}_j}] \le C\delta_j n^{-1}(j\log j)^2(\gamma_j j\log j)^{1/2}\delta_j^{-1/2}\left(\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2}$$
$$\le Cn^{-1}(j\log j)^{3/2}\left(\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2},$$

and

$$\mathsf{E}\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}\sum_{j=1}^{g_n}A_j\mathbb{I}_{\mathcal{E}_j}\Big|X_0\right] \le O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{g_n}(j\log j)^{3/2}\left(\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2}\right).$$

By Cauchy-Schwarz inequality, we have $\sum_{j=1}^{h_n}\gamma_j^{-1} \le h_n^{1/2}\left(\sum_{j=1}^{h_n}\gamma_j^{-2}\right)^{1/2}$, which implies that $n^{-1/4}h_n^{-1/4}\left(\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2} \le \left(n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-2}\right)^{1/4}$. Hence, under Condition (A5), we have

$$\mathsf{E}\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}\sum_{j=1}^{g_n}A_j\mathbb{I}_{\mathcal{E}_j}\Big|X_0\right] = o_{\mathsf{P}}\left(n^{-1/4}h_n^{-1/4}\sum_{j=1}^{g_n}(j\log j)^{3/2}\right),$$

where the upper bound is dominated by $n^{-1/2}h_n^{-1/2}g_n^5(\log g_n)^6$. □

#### 3.10.6.3 Weak convergence of $B_n$

To show the weak convergence of the term $B_n$ in (3.74), the following lemmas about scaling terms are needed.

**Lemma 46.** *Suppose that the conditional variance is given as* $\sigma^2(X) \equiv \sum_{j=1}^{\infty}\gamma_j\rho_j^2\xi_j^2$ *for some* $\{\rho_j\}_{j=1}^{\infty}$ *such that* $\sum_{j=1}^{\infty}\gamma_j\rho_j^2 < \infty$.

*(1) Suppose that the FPC scores* $\{\xi_j\}_{j=1}^{\infty}$ *are independent. Then, we have* $h_n^{-1}s_{h_n}(X_0) \xrightarrow{\mathsf{P}} \sum_{j=1}^{\infty}\gamma_j\rho_j^2$ *as* $n\to\infty$.

*(2) Suppose that* $\xi_j = \xi W_j$ *with* $\mathsf{E}[\xi^4] < \infty$*, where* $\{W_j\}_{j=1}^{\infty}$ *is a sequence of independent random variables with* $\sup_{j\in\mathbb{N}}\mathsf{E}[W_j^4] < \infty$ *(e.g.,* $W_j \overset{iid}{\sim} \mathsf{N}(0,1)$*) and is independent of* $\xi$*. Then, we have* $h_n^{-1}s_{h_n}(X_0) \xrightarrow{\mathsf{P}} \mathsf{E}[\xi^4]\left(\sum_{j=1}^{\infty}\gamma_j\rho_j^2\right)\xi^2$ *as* $n\to\infty$.

*Proof.* Note that the scaling can be expanded as

$$s_{h_n}(X_0) \equiv \langle \Lambda\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle = \left\langle \sum_{j=1}^{h_n} \gamma_j^{-1}\langle X_0, \phi_j\rangle \Lambda\phi_j, \sum_{j=1}^{h_n} \gamma_j^{-1}\langle X_0, \phi_j\rangle\phi_j \right\rangle$$
$$= \sum_{1\le j,j'\le h_n} \gamma_j^{-1}\gamma_{j'}^{-1}\langle X_0, \phi_j\rangle\langle X_0, \phi_{j'}\rangle\langle \Lambda\phi_j, \phi_j'\rangle.$$

Recall that $\Lambda \equiv \mathsf{cov}[X, \varepsilon] = \mathsf{E}[(X\varepsilon)^{\otimes 2}] = \mathsf{E}[X^{\otimes 2}\sigma^2(X)]$, and from the Karhunen–Loève expansion $X = \sum_{j=1}^\infty \gamma_j^{1/2}\xi_j\phi_j$, we have $X^{\otimes 2} = \sum_{j,j'\in\mathbb{N}} \gamma_j^{1/2}\gamma_{j'}^{1/2}\xi_j\xi_{j'}(\phi_j \otimes \phi_{j'})$. This implies that $\langle X^{\otimes 2}\phi_j, \phi_{j'}\rangle = \gamma_j^{1/2}\gamma_{j'}^{1/2}\xi_j\xi_{j'}$ and

$$\langle \Lambda\phi_j, \phi_{j'}\rangle = \mathsf{E}[\sigma^2(X)\langle X^{\otimes 2}\phi_j, \phi_{j'}\rangle] = \gamma_j^{1/2}\gamma_{j'}^{1/2}\mathsf{E}[\xi_j\xi_{j'}\sigma^2(X)].$$

The scaling is then written as $s_{h_n}(X_0) = \sum_{1\le j,j'\le h_n} \xi_j\xi_{j'}\mathsf{E}[\xi_j\xi_{j'}\sigma^2(X)]$. In both cases, we have

$$\mathsf{E}[\xi_j\xi_{j'}\xi_l^2] = \begin{cases} \mathsf{E}[\xi_j^4] & \text{if } j = j' = l \\ \mathsf{E}[\xi_j^2\xi_l^2] & \text{if } j = j \neq l \\ 0 & \text{otherwise,} \end{cases}$$

and hence,

$$s_{h_n}(X_0) = \sum_{1\le j,j'\le h_n} \xi_j\xi_{j'}\mathsf{E}[\xi_j\xi_{j'}\sigma^2(X)] = \sum_{1\le j,j'\le h_n} \xi_j\xi_{j'} \sum_{l=1}^\infty \gamma_l\rho_l^2\mathsf{E}[\xi_j\xi_{j'}\xi_l^2]$$
$$= \sum_{j=1}^{h_n}\sum_{l=1}^\infty \gamma_l\rho_l^2\mathsf{E}[\xi_j^2\xi_l^2]\xi_j^2.$$

The first part is proved as follows. Since $\mathsf{E}[\xi_j^2\xi_l^2] = 1$, we have

$$s_{h_n}(X_0) = \sum_{j=1}^{h_n} \gamma_j\rho_j^2\mathsf{E}[\xi_j^4]\xi_j^2 + \sum_{j=1}^{h_n}\sum_{l\in\mathbb{N}, l\neq j} \gamma_l\rho_l^2\xi_j^2$$
$$= \sum_{j=1}^{h_n} \gamma_j\rho_j^2(\mathsf{E}[\xi_j^4] - 1)\xi_j^2 + \sum_{j=1}^{h_n}\sum_{l=1}^\infty \gamma_l\rho_l^2\xi_j^2.$$

Note that $\mathsf{E}\left[\left|\sum_{j=1}^{h_n} \gamma_j\rho_j^2(\mathsf{E}[\xi_j^4] - 1)\xi_j^2\right|\right] \le \left(\sup_{j\in\mathbb{N}} \mathsf{E}[\xi_j^4] + 1\right) \sum_{j=1}^\infty \gamma_j\rho_j^2 < \infty$ and

$$\mathsf{E}\left[\left(h_n^{-1}\sum_{j=1}^{h_n} \xi_j^2 - 1\right)^2\right] \le h_n^{-2}\sum_{j=1}^{h_n} \mathsf{E}[(\xi_j^2 - 1)^2] \le h_n^{-1}\sup_{j\in\mathbb{N}} \mathsf{E}[\xi_j^4] \le C/h_n \to 0,$$

i.e., $h_n^{-1}\sum_{j=1}^{h_n}\xi_j^2 \xrightarrow{\mathsf{P}} 1$. This implies that

$$\left|h_n^{-1}s_{h_n}(X_0) - \sum_{j=1}^{\infty}\gamma_j\rho_j^2\right| \le h_n^{-1}\left|\sum_{j=1}^{h_n}\gamma_j\rho_j^2(\mathsf{E}[\xi_j^4]-1)\xi_j^2\right| + \left(\sum_{l=1}^{\infty}\gamma_l\rho_l^2\right)\left|h_n^{-1}\sum_{j=1}^{h_n}\xi_j^2 - 1\right|$$
$$= O_{\mathsf{P}}(h_n^{-1}) + o_{\mathsf{P}}(1) = o_{\mathsf{P}}(1),$$

i.e., $h_n^{-1}s_{h_n}(X_0) \xrightarrow{\mathsf{P}} \sum_{j=1}^{\infty}\gamma_j\rho_j^2$ as $n\to\infty$.

The second part is similarly proved as follows. Since $\mathsf{E}[\xi_j^4] = \mathsf{E}[\xi^4]\mathsf{E}[W_j^4]$ and $\mathsf{E}[\xi_j^2\xi_l^2] = \mathsf{E}[\xi^4]\mathsf{E}[W_j^2]\mathsf{E}[W_l^2] = \mathsf{E}[\xi^4]$ under (2), we have

$$s_{h_n}(X_0) = \xi^2\mathsf{E}[\xi^4]\left(\sum_{j=1}^{h_n}\gamma_j\rho_j^2\mathsf{E}[W_j^4]W_j^2 + \sum_{j=1}^{h_n}\sum_{l\in\mathbb{N},l\neq j}\gamma_l\rho_l^2W_j^2\right)$$
$$= \xi^2\mathsf{E}[\xi^4]\left(\sum_{j=1}^{h_n}\gamma_j\rho_j^2(\mathsf{E}[W_j^4]-1)W_j^2 + \sum_{j=1}^{h_n}\sum_{l=1}^{\infty}\gamma_l\rho_l^2W_j^2\right).$$

Note that $\mathsf{E}\left[\left|\sum_{j=1}^{h_n}\gamma_j\rho_j^2(\mathsf{E}[W_j^4]-1)\xi_j^2\right|\right] \le \left(\sup_{j\in\mathbb{N}}\mathsf{E}[W_j^4]+1\right)\sum_{j=1}^{\infty}\gamma_j\rho_j^2 < \infty$ and

$$\mathsf{E}\left[\left(h_n^{-1}\sum_{j=1}^{h_n}W_j^2 - 1\right)^2\right] \le h_n^{-2}\sum_{j=1}^{h_n}\mathsf{E}[(W_j^2-1)^2] \le h_n^{-1}\sup_{j\in\mathbb{N}}\mathsf{E}[W_j^4] \le C/h_n \to 0,$$

i.e., $h_n^{-1}\sum_{j=1}^{h_n}W_j^2 \xrightarrow{\mathsf{P}} 1$. This implies that

$$\left|h_n^{-1}s_{h_n}(X_0) - \xi^2\mathsf{E}[\xi^4]\sum_{j=1}^{\infty}\gamma_j\rho_j^2\right|$$
$$\le \xi^2\mathsf{E}[\xi^4]\left(h_n^{-1}\left|\sum_{j=1}^{h_n}\gamma_j\rho_j^2(\mathsf{E}[W_j^4]-1)\xi_j^2\right| + \left(\sum_{l=1}^{\infty}\gamma_l\rho_l^2\right)\left|h_n^{-1}\sum_{j=1}^{h_n}W_j^2 - 1\right|\right)$$
$$= O_{\mathsf{P}}(1)\{O_{\mathsf{P}}(h_n^{-1}) + o_{\mathsf{P}}(1)\} = o_{\mathsf{P}}(1),$$

i.e., $h_n^{-1}s_{h_n}(X_0) \xrightarrow{\mathsf{P}} \xi^2\mathsf{E}[\xi^4]\sum_{j=1}^{\infty}\gamma_j\rho_j^2$ as $n\to\infty$. □

**Lemma 47.** *Define $r_{h_n}(x) \equiv \langle\Theta\Gamma_{h_n}^{-1}x, \Gamma_{h_n}^{-1}x\rangle$ for $x\in\mathbb{H}$, where $\Theta \equiv \mathsf{E}[\{(X^{\otimes 2}-\Gamma)\beta\}^{\otimes 2}]$. We suppose that the FPCs scores are dependent as $\xi_j = \xi W_j$ with $W_j \overset{iid}{\sim} \mathsf{N}(0,1)$ where $\xi$ is a random variable independent of $\{W_j\}_{j=1}^{\infty}$ and with finite fourth moment $\mathsf{E}[\xi^4] < \infty$. Then, we have $h_n^{-1}r_{h_n}(X_0) \xrightarrow{\mathsf{P}} \mathsf{E}[\xi^4]\|\Gamma^{1/2}\beta\|^2\xi^2$ as $n\to\infty$.*

*Proof.* A direct computation gives $\Theta = \mathsf{E}[\{(X^{\otimes 2} - \Gamma)\beta\}^{\otimes 2}] = \mathsf{E}[(X^{\otimes 2}\beta)^{\otimes 2}] - (\Gamma\beta)^{\otimes 2}$. Note that

$$X^{\otimes 2}\beta = \sum_{j,l\in\mathbb{N}} \gamma_j^{1/2}\gamma_l^{1/2}\xi_j\xi_l\beta_j\phi_l$$

and

$$(X^{\otimes 2}\beta)^{\otimes 2} = \sum_{j,l,j',l'\in\mathbb{N}} \gamma_j^{1/2}\gamma_l^{1/2}\gamma_{j'}^{1/2}\gamma_{l'}^{1/2}\xi_j\xi_l\xi_{j'}\xi_{l'}\beta_j\beta_{j'}(\phi_l\otimes\phi_{l'}).$$

By construction of $\{\xi_j\}$, we have

$$\mathsf{E}[\xi_j\xi_l\xi_{j'}\xi_{l'}] = \mathsf{E}[\xi^4]\mathsf{E}[W_jW_lW_{j'}W_{l'}].$$

Since $\{W_j\}_{j=1}^{\infty}$ is a sequence of independent standard normal random variables, we apply the Isserlis formula to compute the mixed moments:

$$\mathsf{E}[W_jW_lW_{j'}W_{l'}] = \begin{cases} 3 & \text{if } j = j' = l = l' \\ 1 & \text{if } j = j', l = l' \text{ or } j = l, j' = l' \text{ or } j = l', j' = l \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, we have

$$\begin{aligned} \mathsf{E}[(X^{\otimes 2}\beta)^{\otimes 2}]/\mathsf{E}[\xi^4] &= 3\sum_{j=1}^{\infty}\gamma_j^2\beta_j^2\phi_j^{\otimes 2} + \sum_{j,l\in\mathbb{N}}\gamma_j\gamma_l\beta_j^2\phi_l^{\otimes 2} + 2\sum_{j,l\in\mathbb{N}}\gamma_j\gamma_l\beta_j\beta_l(\phi_j\otimes\phi_l) \\ &= 3\sum_{j=1}^{\infty}\gamma_j^2\beta_j^2\phi_j^{\otimes 2} + \|\Gamma^{1/2}\beta\|^2\Gamma + 2(\Gamma\beta)^{\otimes 2}, \end{aligned}$$

which implies that $\Theta = \mathsf{E}[\xi^4]\left\{3\sum_{j=1}^{\infty}\gamma_j^2\beta_j^2\phi_j^{\otimes 2} + \|\Gamma^{1/2}\beta\|^2\Gamma + 2(\Gamma\beta)^{\otimes 2}\right\} - (\Gamma\beta)^{\otimes 2}$. Note that

$$\left\langle \left(\sum_{j=1}^{\infty}\gamma_j^2\beta_j^2\phi_j^{\otimes 2}\right)\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\right\rangle = \sum_{j=1}^{h_n}\beta_j^2\langle X_0,\phi_j\rangle^2,$$

$$\langle \Gamma\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle = \sum_{j=1}^{h_n}\gamma_j^{-1}\langle X_0,\phi_j\rangle^2 = t_{h_n}(X_0),$$

$$\langle (\Gamma\beta)^{\otimes 2}\Gamma_{h_n}^{-1}X_0, \Gamma_{h_n}^{-1}X_0\rangle = \langle \Gamma\beta, \Gamma_{h_n}^{-1}X_0\rangle^2 = \sum_{j=1}^{h_n}\beta_j^2\langle X_0,\phi_j\rangle^2.$$

This implies that

$$r_{h_n}(X_0) = (5\mathsf{E}[\xi^4] - 1)\sum_{j=1}^{h_n} \beta_j^2 \langle X_0, \phi_j\rangle^2 + \mathsf{E}[\xi^4]\|\Gamma^{1/2}\beta\|^2 t_{h_n}(X_0).$$

Since $\sum_{j=1}^{\infty} \gamma_j \beta_j^2 = \|\Gamma^{1/2}\beta\|^2 < \infty$, as $n \to \infty$, we have $\mathsf{E}\left[h_n^{-1}\sum_{j=1}^{h_n} \beta_j^2 \langle X_0, \phi_j\rangle^2\right] = h_n^{-1}\sum_{j=1}^{h_n} \gamma_j\beta_j^2 \le h_n^{-1}\|\Gamma^{1/2}\beta\|^2 \to 0$, which implies that $h_n^{-1}\sum_{j=1}^{h_n} \beta_j^2\langle X_0, \phi_j\rangle^2 \xrightarrow{\mathsf{P}} 0$. By the Law of Large Numbers and Slutsky's theorem, we have $h_n^{-1} t_{h_n}(X_0) = \xi^2 h_n^{-1}\sum_{j=1}^{h_n} W_j^2 \xrightarrow{\mathsf{P}} \xi^2$. Thus,

$$\begin{aligned} h_n r_{h_n}(X_0) =& (5\mathsf{E}[\xi^4] - 1)h_n^{-1}\sum_{j=1}^{h_n} \beta_j^2\langle X_0, \phi_j\rangle^2 + \mathsf{E}[\xi^4]\|\Gamma^{1/2}\beta\|^2 h_n^{-1} t_{h_n}(X_0) \\ &\xrightarrow{\mathsf{P}} \mathsf{E}[\xi^4]\|\Gamma^{1/2}\beta\|^2\xi^2. \end{aligned}$$

□

**Lemma 48.** *Suppose that the FPC scores $\xi_j$ are dependent as $\xi_j = \xi W_j$ with $W_j \overset{iid}{\sim} \mathsf{N}(0,1)$ where $\xi$ is a random variable independent of $\{W_j\}_{j=1}^{\infty}$ with finite eighth moment $\mathsf{E}[\xi^8] < \infty$. Write $\Theta \equiv \mathsf{E}[\{(X^{\otimes 2} - \Gamma)\beta\}^{\otimes 2}]$ and $(\theta_j, \zeta_j)$ for the $j$-th eigenpair of $\Theta$. Then, we have $\sup_{j\in\mathbb{N}} \theta_j^{-2}\mathsf{E}[\langle (X^{\otimes 2} - \Gamma)\beta, \zeta_j\rangle^4] < \infty$*

*Proof.* Recall from Lemma 47 that

$$\begin{aligned} \Theta = \mathsf{E}[\{(X^{\otimes 2} - \Gamma)\beta\}^{\otimes 2}] &= \sum_{j=1}^{\infty} \theta_j \zeta_j^{\otimes 2} \\ &= 3\mathsf{E}[\xi^4]\sum_{j=1}^{\infty} \gamma_j^2\beta_j^2\phi_j^{\otimes 2} + \mathsf{E}[\xi^3]\|\Gamma^{1/2}\beta\|^2\Gamma + (2\mathsf{E}[\xi^4] - 1)(\Gamma\beta)^{\otimes 2} \end{aligned}$$

with $\mathsf{E}[\xi^4] \in [1, \infty)$ so that

$$\theta_j = \langle \Theta\zeta_j, \zeta_j\rangle = 3\mathsf{E}[\xi^4]\sum_{l=1}^{\infty} \gamma_l^2\beta_l^2\langle\phi_l, \zeta_j\rangle^2 + \mathsf{E}[\xi^4]\|\Gamma^{1/2}\beta\|^2\|\Gamma^{1/2}\zeta_j\|^2 + (2\mathsf{E}[\xi^4] - 1)\langle\Gamma\beta, \zeta_j\rangle^2, \quad (3.87)$$

where $\|\Gamma^{1/2}\beta\|^2 = \sum_{l=1}^{\infty} \gamma_l \beta_l^2$, $\|\Gamma^{1/2}\zeta_j\|^2 = \sum_{j=1}^{\infty} \gamma_l \langle \phi_l, \zeta_j\rangle^2$, and $\langle \Gamma\beta, \zeta_j\rangle = \sum_{l=1}^{\infty} \gamma_l \beta_l \langle \phi_l, \zeta_j\rangle$. Note that

$$
\begin{aligned}
Q_j &\equiv \theta_j^{-1/2}\langle (X^{\otimes 2} - \Gamma)\beta, \zeta_j\rangle = \theta_j^{-1/2}[\langle X, \beta\rangle\langle X, \zeta_j\rangle - \langle \Gamma\beta, \zeta_j\rangle] \\
&= \theta_j^{-1/2}\left\{\left(\sum_{l=1}^{\infty} \gamma_l^{1/2}\beta_l \xi_l\right)\left(\sum_{l=1}^{\infty} \gamma_l^{1/2}\xi_l \langle \phi_l, \zeta_j\rangle\right) - \sum_{l=1}^{\infty} \gamma_l \beta_l \langle \phi_l, \zeta_j\rangle\right\} \\
&= \theta_j^{-1/2}\sum_{l=1}^{\infty} \gamma_l \beta_l \langle \phi_l, \zeta_j\rangle(\xi_l^2 - 1) + \theta_j^{-1/2}\sum_{l\neq l'} \gamma_l^{1/2}\gamma_{l'}^{1/2}\beta_{l'}\langle \phi_l, \zeta_j\rangle \xi_l \xi_{l'} \\
&= \theta_j^{-1/2}\xi^2\sum_{l=1}^{\infty} \gamma_l \beta_l \langle \phi_l, \zeta_j\rangle(W_l^2 - 1) + \theta_j^{-1/2}(\xi^2 - 1)\sum_{l=1}^{\infty} \gamma_l \beta_l \langle \phi_l, \zeta_j\rangle \\
&\quad + \theta_j^{-1/2}\sum_{l\neq l'} \gamma_l^{1/2}\gamma_{l'}^{1/2}\beta_{l'}\langle \phi_l, \zeta_j\rangle \xi_l \xi_{l'}.
\end{aligned}
$$

Then,

$$
\begin{aligned}
Q_j^4 \leq C\theta_j^{-2}\Bigg[&\xi^8\left\{\sum_{l=1}^{\infty} \gamma_l \beta_l \langle \phi_l, \zeta_j\rangle(W_l^2 - 1)\right\}^4 \\
&+ \left\{(\xi^2 - 1)\sum_{l=1}^{\infty} \gamma_l \beta_l \langle \phi_l, \zeta_j\rangle\right\}^4 \\
&+ \left(\sum_{l\neq l'} \gamma_l^{1/2}\gamma_{l'}^{1/2}\beta_{l'}\langle \phi_l, \zeta_j\rangle \xi_l \xi_{l'}\right)^4\Bigg].
\end{aligned} \tag{3.88}
$$

We are now showing that $\sup_{j\in\mathbb{N}} Q_j^4 < \infty$ by bounding all three terms on the right-hand side of (3.88).

The expected value of the first term on the right-hand side in (3.88) is bounded as

$$
\begin{aligned}
&\mathsf{E}\left[\left\{\sum_{l=1}^{\infty} \gamma_l \beta_l \langle \phi_l, \zeta_j\rangle(W_l^2 - 1)\right\}^4\right] \\
&= \sum_{l_1,l_2,l_3,l_4} \gamma_{l_1}\gamma_{l_2}\gamma_{l_3}\gamma_{l_4}\beta_{l_1}\beta_{l_2}\beta_{l_3}\beta_{l_4}\langle \phi_{l_1}, \zeta_j\rangle\langle \phi_{l_2}, \zeta_j\rangle\langle \phi_{l_3}, \zeta_j\rangle\langle \phi_{l_4}, \zeta_j\rangle \\
&\qquad\qquad \times \mathsf{E}[(W_{l_1}^2 - 1)(W_{l_2}^2 - 1)(W_{l_3}^2 - 1)(W_{l_4}^2 - 1)] \\
&\leq C\left[\sum_{l=1}^{\infty} \gamma_l^4 \beta_l^4 \langle \phi_l, \zeta_j\rangle^4 + \left(\sum_{l=1}^{\infty} \gamma_l^2 \beta_l^2 \langle \phi_l, \zeta_j\rangle^2\right)^2\right].
\end{aligned}
$$

because the term $\mathsf{E}[(W_{l_1}^2-1)(W_{l_2}^2-1)(W_{l_3}^2-1)(W_{l_4}^2-1)]$ vanishes if there is one index that is not equal to one of the other indices in $\{l_1,l_2,l_3,l_4\}$. Since

$$\sum_{l=1}^{\infty}\gamma_l^4\beta_l^4\langle\phi_l,\zeta_j\rangle^4 \le \left(\sum_{l=1}^{\infty}\gamma_l^2\beta_l^2\langle\phi_l,\zeta_j\rangle^2\right)^2 \le \theta_j^2,$$

we have that

$$\sup_{j\in\mathbb{N}}\theta_j^{-2}\mathsf{E}\left[\left\{\xi^2\sum_{l=1}^{\infty}\gamma_l\beta_l\langle\phi_l,\zeta_j\rangle(W_l^2-1)\right\}^4\right]<\infty.$$

Next, the expected value of the second term on the right-hand side in (3.88) is bounded as

$$\mathsf{E}\left[\left\{(\xi^2-1)\sum_{l=1}^{\infty}\gamma_l\beta_l\langle\phi_l,\zeta_j\rangle\right\}^4\right]\le \mathsf{E}[(\xi^2-1)^4]\langle\Gamma\beta,\zeta_j\rangle^4\le C\theta_j^2.$$

Finally, the expected value of the third term on the right-hand side in (3.88) is written as

$$\begin{aligned}
&\mathsf{E}\left[\left(\sum_{l\neq l'}\gamma_l^{1/2}\gamma_{l'}^{1/2}\beta_{l'}\langle\phi_l,\zeta_j\rangle\xi_l\xi_{l'}\right)^4\right]\\
&=\sum_{l_1\neq l_1',l_2\neq l_2',l_3\neq l_3',l_4\neq l_4'}\gamma_{l_1}^{1/2}\gamma_{l_2}^{1/2}\gamma_{l_3}^{1/2}\gamma_{l_4}^{1/2}\gamma_{l_1'}^{1/2}\gamma_{l_2'}^{1/2}\gamma_{l_3'}^{1/2}\gamma_{l_4'}^{1/2}\beta_{l_1'}\beta_{l_2'}\beta_{l_3'}\beta_{l_4'}\\
&\qquad\times\langle\phi_{l_1},\zeta_j\rangle\langle\phi_{l_2},\zeta_j\rangle\langle\phi_{l_3},\zeta_j\rangle\langle\phi_{l_4},\zeta_j\rangle\\
&\qquad\times\mathsf{E}[\xi_{l_1}\xi_{l_2}\xi_{l_3}\xi_{l_4}\xi_{l_1'}\xi_{l_2'}\xi_{l_3'}\xi_{l_4'}].
\end{aligned}$$

We have the following cases where $u_1,u_2,u_3,u_4\in\{1,2,3,4\}$ denote distinct indices.

1. If $(l_1,l_1')=(l_2,l_2')=(l_3,l_3')=(l_4,l_4')$, then $\mathsf{E}[\xi_{l_1}\xi_{l_2}\xi_{l_3}\xi_{l_4}\xi_{l_1'}\xi_{l_2'}\xi_{l_3'}\xi_{l_4'}]=\mathsf{E}[\xi_{l_1}^4\xi_{l_1'}^4]$.

2. If $(l_{u_1},l_{u_1}')=(l_{u_2},l_{u_2}')=(l_{u_3},l_{u_3}')$ and $(l_{u_4},l_{u_4}')$ are distinct, then

$$\mathsf{E}[\xi_{l_{u_1}}\xi_{l_{u_2}}\xi_{l_{u_3}}\xi_{l_{u_4}}\xi_{l_{u_1}'}\xi_{l_{u_2}'}\xi_{l_{u_3}'}\xi_{l_{u_4}'}]=\mathsf{E}[\xi_{l_{u_1}}^3\xi_{l_{u_1}'}^3\xi_{l_{u_4}}\xi_{l_{u_4}'}]=\begin{cases}\mathsf{E}[\xi_{l_{u_1}}^4\xi_{l_{u_1}'}^4] & \text{if } l_{u_1}=l_{u_4}',l_{u_1}'=l_{u_4}\\ 0 & \text{otherwise.}\end{cases}$$

3. If $(l_{u_1},l_{u_1}')=(l_{u_2},l_{u_2}')$, $(l_{u_3},l_{u_3}')=(l_{u_4},l_{u_4}')$ are distinct, then

$$\mathsf{E}[\xi_{l_{u_1}}\xi_{l_{u_2}}\xi_{l_{u_3}}\xi_{l_{u_4}}\xi_{l_{u_1}'}\xi_{l_{u_2}'}\xi_{l_{u_3}'}\xi_{l_{u_4}'}]=\mathsf{E}[\xi_{l_{u_1}}^2\xi_{l_{u_1}'}^2\xi_{l_{u_3}}^2\xi_{l_{u_3}'}^2]=\begin{cases}\mathsf{E}[\xi_{l_{u_1}}^4\xi_{l_{u_1}'}^4] & \text{if } l_{u_1}=l_{u_3}',l_{u_1}'=l_{u_3}\\ \mathsf{E}[\xi_{l_{u_1}}^2\xi_{l_{u_1}'}^2\xi_{l_{u_3}}^2\xi_{l_{u_3}'}^2] & \text{otherwise.}\end{cases}$$

4. If $(l_{u_1}, l'_{u_1}) = (l_{u_2}, l'_{u_2})$, $(l_{u_3}, l'_{u_3})$, and $(l_{u_4}, l'_{u_4})$ are distinct, then

$$\mathsf{E}[\xi_{l_{u_1}}\xi_{l_{u_2}}\xi_{l_{u_3}}\xi_{l_{u_4}}\xi_{l'_{u_1}}\xi_{l'_{u_2}}\xi_{l'_{u_3}}\xi_{l'_{u_4}}] = \mathsf{E}[\xi^2_{l_{u_1}}\xi^2_{l'_{u_1}}\xi_{l_{u_3}}\xi_{l'_{u_3}}\xi_{l_{u_4}}\xi_{l'_{u_4}}]$$
$$= \begin{cases} \mathsf{E}[\xi^2_{l_{u_1}}\xi^2_{l'_{u_1}}\xi^2_{l_{u_3}}\xi^2_{l'_{u_3}}] & \text{if } l_{u_3} = l'_{u_4}, l'_{u_3} = l_{u_4} \\ \mathsf{E}[\xi^3_{l_{u_1}}\xi^3_{l'_{u_1}}\xi_{l_{u_4}}\xi_{l'_{u_4}}] = 0 & \text{if } l_{u_1} = l'_{u_3}, l'_{u_1} = l_{u_3} \\ \mathsf{E}[\xi^3_{l_{u_1}}\xi^3_{l'_{u_1}}\xi_{l_{u_3}}\xi_{l'_{u_3}}] = 0 & \text{if } l_{u_1} = l'_{u_4}, l'_{u_1} = l_{u_4} \\ 0 & \text{otherwise.} \end{cases}$$

5. If $(l_{u_1}, l'_{u_1})$, $(l_{u_2}, l'_{u_2})$, $(l_{u_3}, l'_{u_3})$, and $(l_{u_4}, l'_{u_4})$ are all distinct, then we have the following three cases that the joint moment $\mathsf{E}[\xi_{l_{u_1}}\xi_{l_{u_2}}\xi_{l_{u_3}}\xi_{l_{u_4}}\xi_{l'_{u_1}}\xi_{l'_{u_2}}\xi_{l'_{u_3}}\xi_{l'_{u_4}}]$ can be non-zero:

   i. exactly two pairs exist among $(l_{u_1}, l_{u_2}, l_{u_3}, l_{u_4})$ and exactly two pairs exist among $(l'_{u_1}, l'_{u_2}, l'_{u_3}, l'_{u_4})$;

   ii. exactly one pair exists among $(l_{u_1}, l_{u_2}, l_{u_3}, l_{u_4})$ and exactly one pair exists among $(l'_{u_1}, l'_{u_2}, l'_{u_3}, l'_{u_4})$;

   iii. no pairs exist among the $l_{u_k}$ but each $l_{u_k}$ must be paired to one and only one $l'_{u_m}$ (no pairing of $l_{u_k}$ to some $l_{u_m}$).

   Only these cases provide non-zero contribution with

$$\mathsf{E}[\xi_{l_{u_1}}\xi_{l_{u_2}}\xi_{l_{u_3}}\xi_{l_{u_4}}\xi_{l'_{u_1}}\xi_{l'_{u_2}}\xi_{l'_{u_3}}\xi_{l'_{u_4}}] = \mathsf{E}[\xi^2_{l_1}\xi^2_{l_2}\xi^2_{l_3}\xi^2_{l_4}] = \mathsf{E}[\xi^8] \neq 0.$$

   Otherwise, the joint moment $\mathsf{E}[\xi_{l_{u_1}}\xi_{l_{u_2}}\xi_{l_{u_3}}\xi_{l_{u_4}}\xi_{l'_{u_1}}\xi_{l'_{u_2}}\xi_{l'_{u_3}}\xi_{l'_{u_4}}]$ is zero.

We are now ready to write down the expansion of the fourth power and arrange the repeated sums by the form of the summation according to the previously derived cases, retaining only summands with a non-zero moment. By bounding each summand by its absolute value and also potentially increase the number of terms to sum over, we obtain

$$
\begin{aligned}
&\mathsf{E}\left[\left(\sum_{l\neq l'}\gamma_l^{1/2}\gamma_{l'}^{1/2}\beta_{l'}\langle\phi_l,\zeta_j\rangle\xi_l\xi_{l'}\right)^4\right]\\
\leq & C\sum_{l\neq l'}\gamma_l^2\gamma_{l'}^2\beta_{l'}^4\langle\phi_l,\zeta_j\rangle^4\\
&+C\sum_{l\neq l'}\gamma_l^2\gamma_{l'}^2|\beta_l\beta_{l'}^3||\langle\phi_l,\zeta_j\rangle|^3|\langle\phi_{l'},\zeta_j\rangle|+C\sum_{l\neq l'}\gamma_l^2\gamma_{l'}^2\beta_l^2\beta_{l'}^2\langle\phi_l,\zeta_j\rangle^2\langle\phi_{l'},\zeta_j\rangle^2\\
&+C\sum_{l_1\neq l_1',l_3\neq l_3'}\gamma_{l_1}\gamma_{l_3}\gamma_{l_1'}\gamma_{l_3'}\beta_{l_1'}^2\beta_{l_3'}^2\langle\phi_{l_1},\zeta_j\rangle^2\langle\phi_{l_3},\zeta_j\rangle^2\\
&+C\sum_{l_1\neq l_1',l_3\neq l_3'}\gamma_{l_1}\gamma_{l_3}\gamma_{l_1'}\gamma_{l_3'}\beta_{l_1'}^2|\beta_{l_3}\beta_{l_3'}|\langle\phi_{l_1},\zeta_j\rangle^2|\langle\phi_{l_3},\zeta_j\rangle||\langle\phi_{l_3'},\zeta_j\rangle|\\
&+C\sum_{l_1,l_2,l_3,l_4\text{ are distinct}}\gamma_{l_1}\gamma_{l_2}\gamma_{l_3}\gamma_{l_4}\beta_{l_1}^2\beta_{l_2}^2\langle\phi_{l_3},\zeta_j\rangle^2\langle\phi_{l_4},\zeta_j\rangle^2\\
&+C\sum_{l_1,l_2,l_3,l_4\text{ are distinct}}\gamma_{l_1}\gamma_{l_2}\gamma_{l_3}\gamma_{l_4}\beta_{l_1}^2|\beta_{l_2}\beta_{l_3}|\langle\phi_{l_2},\zeta_j\rangle^2|\langle\phi_{l_3},\zeta_j\rangle||\langle\phi_{l_4},\zeta_j\rangle|\\
&+C\sum_{l_1,l_2,l_3,l_4\text{ are distinct}}\gamma_{l_1}\gamma_{l_2}\gamma_{l_3}\gamma_{l_4}|\beta_{l_1}\beta_{l_2}\beta_{l_3}\beta_{l_4}||\langle\phi_{l_1},\zeta_j\rangle||\langle\phi_{l_2},\zeta_j\rangle||\langle\phi_{l_3},\zeta_j\rangle||\langle\phi_{l_4},\zeta_j\rangle|\\
\leq & C\left(\sum_{l=1}^{\infty}\gamma_l^2\langle\phi_l,\zeta_j\rangle^4\right)\left(\sum_{l=1}^{\infty}\gamma_l^2\beta_l^4\right)\\
&+C\left(\sum_{l=1}^{\infty}\gamma_l^2|\beta_l||\langle\phi_l,\zeta_j\rangle|^3\right)\left(\sum_{l=1}^{\infty}\gamma_l^2|\beta_l|^3|\langle\phi_l,\zeta_j\rangle|\right)+C\left(\sum_{l=1}^{\infty}\gamma_l^2\beta_l^2\langle\phi_l,\zeta_j\rangle^2\right)^2\\
&+C\left(\sum_{l=1}^{\infty}\gamma_l\langle\phi_l,\zeta_j\rangle^2\right)^2\left(\sum_{l=1}^{\infty}\gamma_l\beta_l^2\right)^2\\
&+C\left(\sum_{l=1}^{\infty}\gamma_l\langle\phi_l,\zeta_j\rangle^2\right)\left(\sum_{l=1}\gamma_l\beta_l^2\right)\left(\sum_{l=1}^{\infty}\gamma_l|\beta_l||\langle\phi_l,\zeta_j\rangle|\right)^2\\
&+C\left(\sum_{1=1}^{\infty}\gamma_l\beta_l^2\right)^2\left(\sum_{l=1}^{\infty}\gamma_l\langle\phi_l,\zeta_j\rangle^2\right)^2\\
&+C\left(\sum_{l=1}^{\infty}\gamma_l\beta_l^2\right)\left(\sum_{l=1}^{\infty}\gamma_l|\beta_l|\langle\phi_l,\zeta_j\rangle^2\right)\left(\sum_{l=1}^{\infty}\gamma_l|\beta_l||\langle\phi_l,\zeta_j\rangle|\right)\left(\sum_{l=1}^{\infty}\gamma_l|\langle\phi_l,\zeta_j\rangle|\right)\\
&+C\left(\sum_{l=1}^{\infty}\gamma_l|\beta_l||\langle\phi_l,\zeta_j\rangle|\right)^4
\end{aligned}
$$

Each term of the last upper bound in the above display is bounded as follows.

(i) The first term is bounded because $\sum_{l=1}^{\infty} \gamma_l^2 \beta_l^4 < \infty$ (due to $\sum_{l=1}^{\infty} \gamma_l \beta_l^2 = \|\Gamma^{1/2}\beta\|^2 < \infty$) and

$$\sum_{l=1}^{\infty} \gamma_l^2 \langle \phi_l, \zeta_j \rangle^4 \leq \left( \sum_{l=1}^{\infty} \gamma_l \langle \phi_l, \zeta_j \rangle^2 \right)^2 \leq C\theta_j^2$$

since all quantities in the sum are positive.

(ii) The third term is bounded because

$$\left( \sum_{l=1}^{\infty} \gamma_l^2 \beta_l^2 \langle \phi_l, \zeta_j \rangle^2 \right)^2 \leq \left( \sum_{l=1}^{\infty} \gamma_l^2 \beta_l^4 \right) \left( \sum_{l=1}^{\infty} \gamma_l^2 \langle \phi_l, \zeta_j \rangle^4 \right) \leq C\theta_j^2$$

by the upper bound for the first term.

(iii) The second term is bounded because

$$\sum_{l=1}^{\infty} \gamma_l^2 |\beta_l| |\langle \phi_l, \zeta_j \rangle|^3 \leq \left( \sum_{l=1}^{\infty} \gamma_l^2 \beta_l^2 \langle \phi_l, \zeta_j \rangle^2 \right)^{1/2} \left( \sum_{l=1}^{\infty} \gamma_l^2 \langle \phi_l, \zeta_j \rangle^4 \right)^{1/2} \leq C\theta_j^{3/2},$$
$$\sum_{l=1}^{\infty} \gamma_l^2 |\beta_l|^3 |\langle \phi_l, \zeta_j \rangle| \leq \left( \sum_{l=1}^{\infty} \gamma_l^2 \beta_l^2 \langle \phi_l, \zeta_j \rangle^2 \right)^{1/2} \left( \sum_{l=1}^{\infty} \gamma_l^2 \beta_l^4 \right)^{1/2} \leq C\theta_j^{1/2},$$

where the upper bounds are obtained from the upper bounds for the first and third terms.

(iv) The fourth and sixth terms are bounded because

$$\begin{aligned} \left( \sum_{l=1}^{\infty} \gamma_l \langle \phi_l, \zeta_j \rangle^2 \right)^2 \left( \sum_{l=1}^{\infty} \gamma_l \beta_l^2 \right)^2 &= \|\Gamma^{1/2}\beta\|^4 \left( \sum_{l=1}^{\infty} \gamma_l \langle \phi_l, \zeta_j \rangle^2 \right)^2 \\ &\leq \mathsf{E}[\xi^4]^2 \|\Gamma^{1/2}\beta\|^4 \left( \sum_{l=1}^{\infty} \gamma_l \langle \phi_l, \zeta_j \rangle^2 \right)^2 \leq \theta_j^2, \end{aligned}$$

where the inequality is due to (3.87).

(v) The fifth term is bounded because

$$\left( \sum_{l=1}^{\infty} \gamma_l |\beta_l| |\langle \phi_l, \zeta_j \rangle| \right)^2 \leq \left( \sum_{l=1}^{\infty} \gamma_l \beta_l^2 \right) \left( \sum_{l=1}^{\infty} \gamma_l \langle \phi_l, \zeta_j \rangle^2 \right) = \|\Gamma^{1/2}\beta\|^2 \|\Gamma^{1/2}\zeta_j\|^2$$

and hence

$$\begin{aligned} \left( \sum_{l=1}^{\infty} \gamma_l \langle \phi_l, \zeta_j \rangle^2 \right) \left( \sum_{l=1} \gamma_l \beta_l^2 \right) \left( \sum_{l=1}^{\infty} \gamma_l |\beta_l| \langle \phi_l, \zeta_j \rangle \right)^2 &= \|\Gamma^{1/2}\beta\|^4 \|\Gamma^{1/2}\zeta_j\|^4 \\ &\leq \mathsf{E}[\xi^4]^2 \|\Gamma^{1/2}\beta\|^4 \|\Gamma^{1/2}\zeta_j\|^4 \leq \theta_j^2. \end{aligned}$$

(vi) To bound the seventh term, note that

$$\left(\sum_{l=1}^{\infty}\gamma_l|\beta_l|\langle\phi_l,\zeta_j\rangle^2\right)^2 \le \left(\sum_{l=1}^{\infty}\beta_l^2\right)\left(\sum_{l=1}^{\infty}\gamma_l^2\langle\phi_l,\zeta_j\rangle^4\right) \le C\theta_j^2$$

due to the upper bound for the first term. We also have

$$\left(\sum_{l=1}^{\infty}\gamma_l|\beta_l||\langle\phi_l,\zeta_j\rangle|\right)^2 \le \left(\sum_{l=1}^{\infty}\gamma_l\beta_l^2\right)\left(\sum_{l=1}^{\infty}\gamma_l\langle\phi_l,\zeta_j\rangle^2\right) = \|\Gamma^{1/2}\beta\|^2\|\Gamma^{1/2}\zeta_j\|^2$$
$$\le \mathsf{E}[\xi^4]\|\Gamma^{1/2}\beta\|^2\|\Gamma^{1/2}\zeta_j\|^2 \le \theta_j.$$

It follows that

$$\left(\sum_{l=1}^{\infty}\gamma_l|\langle\phi_l,\zeta_j\rangle|\right)^2 \le \left(\sum_{l=1}^{\infty}\gamma_l\right)\left(\sum_{l=1}^{\infty}\gamma_l\langle\phi_l,\zeta_j\rangle^2\right) = \mathrm{tr}(\Gamma)\|\Gamma^{1/2}\zeta_j\|^2 \le C\theta_j.$$

Then, the seventh term is bounded as

$$\left(\sum_{l=1}^{\infty}\gamma_l\beta_l^2\right)\left(\sum_{l=1}^{\infty}\gamma_l|\beta_l|\langle\phi_l,\zeta_j\rangle^2\right)\left(\sum_{l=1}^{\infty}\gamma_l|\beta_l||\langle\phi_l,\zeta_j\rangle|\right)\left(\sum_{l=1}^{\infty}\gamma_l|\langle\phi_l,\zeta_j\rangle|\right) \le C\theta_j^2.$$

(vii) The last term is bounded as above because

$$\left(\sum_{l=1}^{\infty}\gamma_l|\beta_l||\langle\phi_l,\zeta_j\rangle|\right)^4 \le \left(\sum_{l=1}^{\infty}\gamma_l\beta_l^2\right)^2\left(\sum_{l=1}^{\infty}\gamma_l\langle\phi_l,\zeta_j\rangle^2\right)^2 = \|\Gamma^{1/2}\beta\|^4\|\Gamma^{1/2}\zeta_j\|^4$$
$$\le \mathsf{E}[\xi^4]^2\|\Gamma^{1/2}\beta\|^4\|\Gamma^{1/2}\zeta_j\|^4 \le \theta_j^2.$$

We thus conclude that $\sup_{j\in\mathbb{N}} Q_j^4 < \infty$. □

**Proposition 29.** *We have that*

$$\mathsf{E}[(B_{1n}+B_{2n})^2|X_0] \le O_{\mathsf{P}}\left(1-\frac{g_n}{h_n}\right) + \frac{s_{h_n}((I-\Pi_{g_n})X_0)}{s_{h_n}(X_0)} + O_{\mathsf{P}}(n^{-1}).$$

*Thus, as $n\to\infty$, if $h_n/g_n \to \tau = 1$ and $\frac{s_{h_n}((I-\Pi_{g_n})X_0)}{s_{h_n}(X_0)} \xrightarrow{\mathsf{P}} 0$ (which is the case under the assumptions in Lemma 46 along with $h_n/g_n\to\tau=1$), then $\mathsf{E}[(B_{1n}+B_{2n})^2|X_0] \xrightarrow{\mathsf{P}} 0$.*

*Proof.* Note that

$$\langle\Gamma_{h_n}^{-1}(I-\Pi_{g_n})(\hat{\Gamma}_n-\Gamma)\beta, X_0\rangle = \sum_{j>g_n}^{h_n}\sum_{l=1}^{\infty}\gamma_j^{-1}\beta_l\langle X_0,\phi_j\rangle\langle(\hat{\Gamma}_n-\Gamma)\phi_l,\phi_j\rangle$$
$$= \sum_{j>g_n}^{h_n}\sum_{l=1}^{\infty}\gamma_j^{-1/2}\gamma_l^{1/2}\beta_l\langle X_0,\phi_j\rangle\frac{\langle(\hat{\Gamma}_n-\Gamma)\phi_l,\phi_j\rangle}{\gamma_l^{1/2}\gamma_j^{1/2}}$$

Since $\{\langle X_0, \phi_j\rangle\}_{j=1}^{\infty}$ are uncorrelated, denoting $H_{lj} \equiv \frac{\langle(\hat{\Gamma}_n-\Gamma)\phi_l,\phi_j\rangle}{\gamma_l^{1/2}\gamma_j^{1/2}}$, we have

$$\begin{aligned}\mathsf{E}[\langle\Gamma_{h_n}^{-1}(I-\Pi_{g_n})(\hat{\Gamma}_n-\Gamma)\beta, X_0\rangle^2] &= \sum_{j>g_n}^{h_n} \gamma_j^{-1}\mathsf{E}[\langle X_0,\phi_j\rangle^2]\mathsf{E}\left[\left(\sum_{l=1}^{\infty}\gamma_l^{1/2}\beta_l H_{lj}\right)^2\right] \\ &= \sum_{j>g_n}^{h_n} \mathsf{E}\left[\left(\sum_{l=1}^{\infty}\gamma_l^{1/2}\beta_l H_{lj}\right)^2\right].\end{aligned}$$

By Lemma 13,

$$\begin{aligned}\mathsf{E}\left[\left(\sum_{l=1}^{\infty}\gamma_l^{1/2}\beta_l H_{lj}\right)^2\right] &= \sum_{l=1}^{\infty}\gamma_l\beta_l^2\mathsf{E}[H_{lj}^2] + \sum_{l\neq l'}\gamma_l^{1/2}\gamma_{l'}^{1/2}\beta_l\beta_{l'}\mathsf{E}[H_{lj}H_{l'j}] \\ &\leq \sum_{l=1}^{\infty}\gamma_l\beta_l^2\mathsf{E}[H_{lj}^2] + \sum_{l\neq l'}\gamma_l^{1/2}\gamma_{l'}^{1/2}|\beta_l||\beta_{l'}|\mathsf{E}[|H_{lj}||H_{l'j}|] \\ &\leq \sum_{l=1}^{\infty}\gamma_l\beta_l^2\mathsf{E}[H_{lj}^2] + \sum_{l\neq l'}\gamma_l^{1/2}\gamma_{l'}^{1/2}|\beta_l||\beta_{l'}|\mathsf{E}[H_{lj}^2]^{1/2}\mathsf{E}[H_{l'j}^2]^{1/2} \\ &\leq \frac{C}{n}\left(\sum_{l=1}^{\infty}\gamma_l\beta_l^2 + \sum_{l\neq l'}\gamma_l^{1/2}\gamma_{l'}^{1/2}|\beta_l||\beta_{l'}|\right) = \frac{C}{n}\left(\sum_{l=1}^{\infty}\gamma_l^{1/2}|\beta_l|\right)^{1/2}.\end{aligned}$$

This implies that

$$\frac{n}{h_n}\mathsf{E}[\langle\Gamma_{h_n}^{-1}(I-\Pi_{g_n})(\hat{\Gamma}_n-\Gamma)\beta, X_0\rangle^2] \leq C\left(1-\frac{g_n}{h_n}\right),$$

and hence,

$$\mathsf{E}\left[\frac{n}{s_{h_n}(X_0)}\langle\Gamma_{h_n}^{-1}(I-\Pi_{g_n})(\hat{\Gamma}_n-\Gamma)\beta, X_0\rangle^2\Big|X_0\right] \leq O_{\mathsf{P}}\left(1-\frac{g_n}{h_n}\right).$$

Note that

$$\langle\Gamma_{h_n}^{-1}(I-\Pi_{g_n})U_n, X_0\rangle = \langle\Gamma_{h_n}^{-1}U_n, (I-\Pi_{g_n})X_0\rangle.$$

Write $\tilde{U}_n \equiv n^{-1}\sum_{i=1}^n X_i\varepsilon_i$ so that $U_n = \tilde{U}_n - \bar{X}\bar{\varepsilon}$. Then, by using the same technique used in Proposition 15, we have

$$\mathsf{E}\left[\frac{n}{s_{h_n}(X_0)}\langle\Gamma_{h_n}^{-1}\bar{X}\bar{\varepsilon}, (I-\Pi_{g_n})X_0\rangle^2\Big|X_0\right] = O_{\mathsf{P}}(n^{-1})$$

and

$$\mathsf{E}\left[\frac{n}{s_{h_n}(X_0)}\langle\Gamma_{h_n}^{-1}\tilde{U}_n,(I-\Pi_{g_n})X_0\rangle^2\Big|X_0\right]=\frac{s_{h_n}((I-\Pi_{g_n})X_0)}{s_{h_n}(X_0)}.$$

We finally obtain

$$\frac{s_{h_n}((I-\Pi_{g_n})X_0)}{s_{h_n}(X_0)}=1-\frac{s_{g_n}(X_0)}{s_{h_n}(X_0)}=1-\frac{g_n}{h_n}\frac{g_n^{-1}s_{g_n}(X_0)}{h_n^{-1}s_{h_n}(X_0)}\xrightarrow{\mathsf{P}}1-\tau^{-1}=0$$

under the assumptions in Lemma 46 along with $h_n/g_n\to\tau=1$. □

**Proposition 30.** *Write $\Theta\equiv\mathsf{E}[\{(X^{\otimes 2}-\Gamma)\beta\}^{\otimes 2}]$ and $(\theta_j,\zeta_j)$ for the $j$-th eigenpair of $\Theta$. Define $r_{h_n}(x)\equiv\langle\Theta\Gamma_{h_n}^{-1}x,\Gamma_{h_n}^{-1}x\rangle$ for $x\in\mathbb{H}$. Suppose the following.*

1. *As $n\to\infty$, $h_n/g_n\to\tau\in(1,\infty)$, $n^{-1}(h_n-g_n)^2\to 0$, and $n^{-1/2}(h_n-g_n)^{-1}\sum_{j>g_n}^{h_n}\gamma_j^{-1}\to 0$, where the last condition is implied by Condition (A5).*

2. *$\sup_{j\in\mathbb{N}}\gamma_j^{-1}\|\Lambda^{1/2}\phi_j\|^2<\infty$, which is implied by $\mathsf{E}[\varepsilon^4]<\infty$ along with Condition (A2).*

3. *$(h_n-g_n)\{r_{h_n}((I-\Pi_{g_n})X_0)+s_{h_n}((I-\Pi_{g_n})X_0)\}^{-1}=O_{\mathsf{P}}(1)$, which holds under either assumptions in Lemma 47.*

4. *$\sup_{j\in\mathbb{N}}\theta_j^{-2}\mathsf{E}[\langle(X^{\otimes 2}-\Gamma)\beta,\zeta_j\rangle^4]<\infty$ and $\mathsf{E}[\|X\|^8]<\infty$.*

*Then, as $n\to\infty$, we have*

$$\sup_{y\in\mathbb{R}}\left|\mathsf{P}\left(\sqrt{\frac{n}{q_n(X_0)}}\langle\Gamma_{h_n}^{-1}(I-\Pi_{g_n})(U_n+(\hat{\Gamma}_n-\Gamma)\beta),X_0\rangle\le y\Big|X_0\right)-\Phi(y)\right|\xrightarrow{\mathsf{P}}0,$$

*where $q_n(X_0)\equiv r_{h_n}((I-\Pi_{g_n})X_0)+s_{h_n}((I-\Pi_{g_n})X_0)$.*

*Proof.* Write $V_{0,n}\equiv\sum_{j>g_n}^{h_n}\gamma_j^{-1}\langle X_0,\phi_j\rangle\phi_j=(I-\Pi_{g_n})\Gamma_{h_n}^{-1}X_0=\Gamma_{h_n}^{-1}(I-\Pi_{g_n})X_0$, $Z_{i,1n}\equiv\langle(X_i^{\otimes 2}-\Gamma)\beta,V_{0,n}\rangle$, $Z_{i,2n}\equiv\langle X_i\varepsilon_i,V_{0,n}\rangle$, and $Z_{i,n}=Z_{i,1n}+Z_{i,2n}$ so that

$$\begin{aligned}&\langle\Gamma_{h_n}^{-1}(I-\Pi_{g_n})(U_n+(\hat{\Gamma}_n-\Gamma)\beta),X_0\rangle\\=&n^{-1}\sum_{i=1}^{n}Z_{i,n}-\langle\bar{X}^{\otimes 2}\beta,V_{0,n}\rangle-\langle\bar{X}\bar{\varepsilon},V_{0,n}\rangle^2.\end{aligned}\tag{3.89}$$

Note that

$$
\begin{aligned}
&\mathsf{E}\left[\sqrt{\frac{n}{q_n(X_0)}}|\langle \bar{X}^{\otimes 2}\beta, V_{0,n}\rangle|\Big|X_0\right] \\
\leq&[h_n\{r_{h_n}((I-\Pi_{g_n})X_0)+s_{h_n}((I-\Pi_{g_n})X_0)\}^{-1}]^{1/2}n^{1/2}h_n^{-1/2}\mathsf{E}[\|\bar{X}\|^2]\|\beta\|\|V_{0,n}\| \\
=&O_{\mathsf{P}}(1)n^{1/2}h_n^{-1/2}O(n^{-1})\|\beta\|O_{\mathsf{P}}\left(\left(\sum_{j>g_n}^{h_n}\gamma_j^{-1}\right)^{1/2}\right) \\
=&O_{\mathsf{P}}\left(\left(n^{-1}(h_n-g_n)^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2}\right).
\end{aligned}
$$

By the first condition, the second term in (3.89) is ignorable.

The contribution of the third term in (3.89) is also negligible as follows. Note that

$$
\langle \bar{X}\bar{\varepsilon}, V_{0,n}\rangle^2 \leq n^{-4}2\left\langle\sum_{i=1}^{n}X_i\varepsilon_i, V_{0,n}\right\rangle^2 + n^{-4}2\left\langle\sum_{i\neq i'}X_i\varepsilon_{i'}, V_{0,n}\right\rangle^2.
$$

Due to the independence of the sample $\{(X_i,Y_i)\}_{i=1}^n$ and $\mathsf{E}[X\varepsilon]=0$, we have that

$$
\begin{aligned}
n^{-1}\mathsf{E}\left[\left\langle\sum_{i=1}^{n}X_i\varepsilon_i, V_{0,n}\right\rangle^2\Big|X_0\right] &= \mathsf{E}[\langle X_i\varepsilon_i, \Gamma_{h_n}^{-1}X_0\rangle^2|X_0] = \mathsf{E}[\langle (X_i\varepsilon_i)^{\otimes 2}V_{0,n}, V_{0,n}\rangle^2|X_0] \\
&= \langle \Lambda V_{0,n}, V_{0,n}\rangle = s_{h_n}((I-\Pi_{g_n})X_0)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathsf{E}\left[\left\langle\sum_{i\neq i'}X_i\varepsilon_{i'}, V_{0,n}\right\rangle^2\Big|X_0\right] &= (n^2-n)\mathsf{E}[\langle X_i\varepsilon_{i'}, V_{0,n}\rangle^2|X_0] \\
&= (n^2-n)\mathsf{E}[\langle \varepsilon_{i'}^2X_i^{\otimes 2}V_{0,n}, V_{0,n}\rangle^2|X_0] \\
&= (n^2-n)\mathsf{E}[\varepsilon^2]\langle \Gamma V_{0,n}, V_{0,n}\rangle \\
&= (n^2-n)\mathsf{E}[\varepsilon^2]\{t_{h_n}(X_0)-t_{g_n}(X_0)\}.
\end{aligned}
$$

Since $\mathsf{E}[s_{h_n}((I-\Pi_{g_n})X_0)] \le C(h_n - g_n)$ due to the assumption $\sup_{j\in\mathbb{N}} \gamma_j^{-1}\|\Lambda^{1/2}\phi_j\|^2 < \infty$ and $\mathsf{E}[t_{h_n}(X_0)] = h_n$, it holds that

$$
\begin{aligned}
&\mathsf{E}\left[\frac{n}{q_n(X_0)}\langle \Gamma_{h_n}^{-1}\bar{X}\bar{\varepsilon}, X_0\rangle^2 \middle| X_0\right] \\
\le & O_{\mathsf{P}}\left(\frac{n}{h_n - g_n}\right)\left[\frac{s_{h_n}((I-\Pi_{g_n})X_0)}{n^3} + \frac{(n^2-n)\mathsf{E}[\varepsilon^2]\{t_{h_n}(X_0) - t_{g_n}(X_0)\}}{n^4}\right] \\
= & O_{\mathsf{P}}\left(\frac{n}{h_n - g_n}\right)\left\{O_{\mathsf{P}}\left(\frac{h_n - g_n}{n^3}\right) + O_{\mathsf{P}}\left(\frac{h_n - g_n}{n^2}\right)\right\} \\
= & O_{\mathsf{P}}(n^{-2})
\end{aligned}
$$

by the assumption (2).

Before showing the weak convergence of the first term in (3.89), we claim the following:

$$
\mathsf{E}\left[\left\| n^{-1}\sum_{i=1}^{n}[\{(X_i^{\otimes 2} - \Gamma)\beta\} \otimes (X_i\varepsilon_i)]\right\|_\infty^2\right] = O_{\mathsf{P}}(n^{-1}); \tag{3.90}
$$

$$
\mathsf{E}\left[\left\| n^{-1}\sum_{i=1}^{n}[(X_i\varepsilon_i) \otimes \{(X_i^{\otimes 2} - \Gamma)\beta\}]\right\|_\infty^2\right] = O_{\mathsf{P}}(n^{-1}). \tag{3.91}
$$

Set $L_i \equiv \{(X_i^{\otimes 2} - \Gamma)\beta\} \otimes (X_i\varepsilon_i)$. We then observe that $\mathsf{E}[L_i] = 0$ and

$$
\begin{aligned}
\mathsf{E}[\|L_i\|_\infty^2] &\le \mathsf{E}[\|X_i^{\otimes 2} - \Gamma\|_\infty^2 \|\beta\|^2 \|X_i\varepsilon_i\|^2] \\
&\le \|\beta\|^2 \mathsf{E}[(\|X_i\|^2 + \mathsf{E}[\|X_i\|^2])^2 \|X_i\varepsilon_i\|^2] \\
&\le 2\|\beta\|^2 \mathsf{E}[(\|X_i\|^4 + \mathsf{E}[\|X_i\|^2]^2)\|X_i\varepsilon_i\|^2] \\
&\le 2\|\beta\|^2 \mathsf{E}[(\|X_i\|^4 + \mathsf{E}[\|X_i\|^4])\|X_i\varepsilon_i\|^2] \\
&\le 2\|\beta\|^2 \mathsf{E}[(\|X_i\|^4 + \mathsf{E}[\|X_i\|^4])^2]^{1/2} \mathsf{E}[\|X_i\varepsilon_i\|^4]^{1/2} \\
&\le 2\sqrt{2}\|\beta\|^2 \mathsf{E}[(\|X_i\|^8 + \mathsf{E}[\|X_i\|^4]^2)]^{1/2} \mathsf{E}[\|X_i\varepsilon_i\|^4]^{1/2} \\
&\le 4\|\beta\|^2 \mathsf{E}[\|X_i\|^8]^{1/2} \mathsf{E}[\|X_i\varepsilon_i\|^4]^{1/2} \\
&< \infty
\end{aligned}
$$

by the assumption $\mathsf{E}[\|X\|^8] < \infty$ and Condition (A7). Since $L_i$'s are iid, Equation (3.90) follows from Theorem 2.5 of [32]. Equation (3.91) can be derived at the same way.

To show the weak convergence of the first term in (3.89), we will derive the Lindberg condition. Define $\mathcal{L} = v_n^{-2} \sum_{i=1}^n \mathsf{E}^{X_0}[Z_{i,n}^2 \mathbb{I}(|Z_{i,n}| > \eta v_n)]$ for $\eta > 0$, where $v_n^2 = \sum_{i=1}^n \mathsf{E}^{X_0}[Z_{i,n}^2]$. The Lindeberg condition is then proved by showing the following propositions: as $n \to \infty$,

$$\frac{n^{-1} v_n^2}{q_n(X_0)} \xrightarrow{\mathsf{P}} 1; \tag{3.92}$$

$$\mathsf{E}^{X_0}\left[v_n^{-1} \max_{1 \le i \le n} |Z_{i,n}|^4\right] \xrightarrow{\mathsf{P}} 0; \tag{3.93}$$

$$\mathsf{E}^{X_0}\left[\left|\frac{n^{-1} Z_{i,n}^2}{n^{-1} v_n^2} - 1\right|^2\right] \xrightarrow{\mathsf{P}} 0. \tag{3.94}$$

We then get the desired result by the same argument as the one in Proposition 15 and Proposition 16.

To derive the convergence in (3.92), note that

$$\begin{aligned}
&\left|\frac{n^{-1} v_n^2}{q_n(X_0)} - 1\right| \\
=&\frac{\left|\left\langle n^{-1} \sum_{i=1}^n [\{(X_i^{\otimes 2} - \Gamma)\beta\} \otimes (X_i \varepsilon_i)] V_{0,n}, V_{0,n}\right\rangle + \left\langle n^{-1} \sum_{i=1}^n [(X_i \varepsilon_i) \otimes \{(X_i^{\otimes 2} - \Gamma)\beta\}] V_{0,n}, V_{0,n}\right\rangle\right|}{q_n(X_0)} \\
\le& O_{\mathsf{P}}((h_n - g_n)^{-1}) \Bigg( \left\| n^{-1} \sum_{i=1}^n [\{(X_i^{\otimes 2} - \Gamma)\beta\} \otimes (X_i \varepsilon_i)] \right\|_\infty \\
&\qquad + \left\| n^{-1} \sum_{i=1}^n [(X_i \varepsilon_i) \otimes \{(X_i^{\otimes 2} - \Gamma)\beta\}] \right\|_\infty \Bigg) \|V_{0,n}\|^2 \\
=& O_{\mathsf{P}}\left( n^{-1/2} (h_n - g_n)^{-1} \sum_{j > g_n}^{h_n} \gamma_j^{-1} \right).
\end{aligned}$$

This approaches to zero by the the assumption (1).

To show the convergence in (3.93), note that

$$|Z_{i,n}| \le \langle \Theta_{g_n,h_n}^{-1/2} (X_i^{\otimes 2} - \Gamma)\beta, \Theta_{g_n,h_n} V_{0,n}\rangle + \langle \Lambda_{g_n,h_n}^{-1/2} X_i \varepsilon_i, \Lambda_{g_n,h_n}^{1/2} V_{0,n}\rangle$$

where $\Theta_{g_n,h_n} \equiv \sum_{j>g_n}^{h_n} \theta_j(\zeta_j \otimes \zeta_j)$ and $\Lambda_{g_n,h_n} \equiv \sum_{j>g_n}^{h_n} \lambda_j(\psi_j \otimes \psi_j)$. Then,

$$
\begin{aligned}
& v_n^{-1} \max_{1\le i\le n} |Z_{i,n}| \\
\le & \frac{n^{-1/2}\max_{1\le i\le n} \|\Theta_{g_n,h_n}^{-1/2}(X_i^{\otimes 2}-\Gamma)\beta\|\|\Theta_{g_n,h_n}V_{0,n}\| + n^{-1/2}\max_{1\le i\le n}\|\Lambda_{g_n,h_n}^{-1/2}X_i\varepsilon_i\|\|\Lambda_{g_n,h_n}^{1/2}V_{0,n}\|}{\{q_n(X_0)\}^{1/2}} \\
& \times \left(\frac{q_n(X_0)}{n^{-1}v_n^2}\right)^{1/2} \\
= & \Bigg[ n^{-1/2}\max_{1\le i\le n} \|\Theta_{g_n,h_n}^{-1/2}(X_i^{\otimes 2}-\Gamma)\beta\| \left\{\frac{r_{h_n}((I-\Pi_{g_n})X_0)}{q_n(X_0)}\right\}^{1/2} \\
& + n^{-1/2}\max_{1\le i\le n}\|\Lambda_{g_n,h_n}^{-1/2}X_i\varepsilon_i\| \left\{\frac{s_{h_n}((I-\Pi_{g_n})X_0)}{q_n(X_0)}\right\}^{1/2} \Bigg] \\
& \times \left(\frac{q_n(X_0)}{n^{-1}v_n^2}\right)^{1/2} \\
= & \left\{ n^{-1/2}\max_{1\le i\le n} \|\Theta_{g_n,h_n}^{-1/2}(X_i^{\otimes 2}-\Gamma)\beta\| + n^{-1/2}\max_{1\le i\le n}\|\Lambda_{g_n,h_n}^{-1/2}X_i\varepsilon_i\| \right\} \left(\frac{q_n(X_0)}{n^{-1}v_n^2}\right)^{1/2}.
\end{aligned}
$$

From Jensen's inequality, we also see that

$$
\|\Theta_{g_n,h_n}^{-1/2}(X_i^{\otimes 2}-\Gamma)\beta\|^2 = \sum_{j>g_n}^{h_n} \theta_j^{-1}\langle (X_i^{\otimes 2}-\Gamma)\beta, \zeta_j\rangle^2 \le \sqrt{(h_n-g_n)\sum_{j>g_n}^{h_n}\theta_j^{-2}\langle (X_i^{\otimes 2}-\Gamma)\beta,\zeta_j\rangle^4},
$$

which implies that

$$
\mathsf{E}\left[\max_{1\le i\le n}\|\Theta_{g_n,h_n}^{-1/2}(X_i^{\otimes 2}-\Gamma)\beta\|^4\right] \le (h_n-g_n)\sum_{i=1}^{n}\sum_{j>g_n}^{h_n}\theta_j^{-2}\mathsf{E}[\langle (X_i^{\otimes 2}-\Gamma)\beta,\zeta_j\rangle^4] \le Cn(h_n-g_n)^2.
$$

At the same way, we obtain

$$
\mathsf{E}\left[\max_{1\le i\le n}\|\Lambda_{g_n,h_n}^{-1/2}X_i\varepsilon_i\|^4\right] \le Cn(h_n-g_n)^2.
$$

We therefore have that

$$
\mathsf{E}^{X_0}\left[\left(v_n^{-1}\max_{1\le i\le n}|Z_{i,n}|\right)^4\right] = O_{\mathsf{P}}(n^{-1}(h_n-g_n)^2)\{1+o_{\mathsf{P}}(1)\},
$$

which converges to zero when $n^{-1}(h_n-g_n)^2 \to 0$.

For the convergence in (3.94), by the assumption (2), we observe that

$$\begin{aligned}
&\left|\frac{n^{-1}\sum_{i=1}^n Z_{i,n}^2}{n^{-1}v_n^2}-1\right| \\
=&\left|\frac{\langle\{(\tilde{\Theta}_n-\Theta)+(\tilde{\Lambda}_n-\Lambda)+Q_{1n}+Q_{2n}\}V_{0,n},V_{0,n}\rangle}{q_n(X_0)}\right| \\
\le& O_{\mathsf{P}}((h_n-g_n)^{-1})(\|\tilde{\Theta}_n-\Theta\|_\infty+\|\tilde{\Lambda}_n-\Lambda\|_\infty+\|Q_{1n}\|_\infty+\|Q_{2n}\|_\infty)\|V_{0,n}\|^2,
\end{aligned}$$

where $\tilde{\Theta}_n \equiv n^{-1}\sum_{i=1}^n\{(X_i^{\otimes 2}-\Gamma)\beta\}^{\otimes 2}$, $\tilde{\Lambda}_n \equiv n^{-1}\sum_{i=1}^n (X_i\varepsilon_i)^{\otimes 2}$, and

$$Q_{1n} \equiv \left\langle n^{-1}\sum_{i=1}^n [\{(X_i^{\otimes 2}-\Gamma)\beta\}\otimes(X_i\varepsilon_i)]V_{0,n},V_{0,n}\right\rangle,$$
$$Q_{2n} \equiv \left\langle n^{-1}\sum_{i=1}^n [(X_i\varepsilon_i)\otimes\{(X_i^{\otimes 2}-\Gamma)\beta\}]V_{0,n},V_{0,n}\right\rangle.$$

Since both $\{(X_i^{\otimes 2}-\Gamma)\beta\}$ and $\{(X_i\varepsilon_i)^{\otimes 2}\}$ are sequences of iid random elements with finite second moments, we have $\mathsf{E}[\|\tilde{\Theta}_n-\Theta\|_\infty^2]=O(n^{-1})$ and $\|\tilde{\Lambda}_n-\Lambda\|_\infty=O(n^{-1})$. By Equations (3.90)-(3.91), we also have $\|Q_{1n}\|_\infty=O(n^{-1})$ and $\|Q_{2n}\|_\infty=O(n^{-1})$. This implies that

$$\left|\frac{n^{-1}\sum_{i=1}^n Z_{i,n}^2}{n^{-1}v_n^2}-1\right| = O_{\mathsf{P}}\left(n^{-1/2}(h_n-g_n)^{-1}\sum_{j>g_n}^{h_n}\gamma_j^{-1}\right).$$

Thus, this converges to zero by the assumption (1). □

**Proposition 31.** *Write $\Theta \equiv \mathsf{E}[\{(X^{\otimes 2}-\Gamma)\beta\}^{\otimes 2}]$ and $(\theta_j,\zeta_j)$ for the $j$-th eigenpair of $\Theta$. We suppose the following:*

1. *as $n\to\infty$, $h_n/g_n\to\tau\in(1,\infty)$, $n^{-1}(h_n-g_n)^2\to 0$, and $n^{-1/2}(h_n-g_n)^{-1}\sum_{j>g_n}^{h_n}\gamma_j^{-1}\to 0$;*
2. *the FPC scores $\xi_j$ are dependent as $\xi_j=\xi W_j$ with $W_j \overset{iid}{\sim} \mathsf{N}(0,1)$ where $\xi$ is a random variable independent of $\{W_j\}_{j=1}^\infty$ with finite eighth moment $\mathsf{E}[\xi^8]<\infty$;*
3. *the conditional variance is given as $\sigma^2(X)\equiv\sum_{j=1}^\infty\gamma_j\rho_j^2\xi_j^2$ for some $\{\rho_j\}_{j=1}^\infty$ such that $\sum_{j=1}^\infty\gamma_j\rho_j^2<\infty$.*

*Then, as $n\to\infty$, we have*

$$\sup_{y\in\mathbb{R}}|\mathsf{P}(B_{1n}+B_{2n}\le y|X_0)-\Phi(y/\sigma_{\text{cons}}(\tau))| \xrightarrow{\mathsf{P}} 0$$

*where* $\sigma^2_{\text{cons}}(\tau) \equiv (1-\tau^{-1})\left(\|\Gamma^{1/2}\beta\|^2 \big/ \sum_{j=1}^\infty \gamma_j\rho_j^2 + 1\right)$.

*Proof.* Define $r_{h_n}(x) \equiv \langle \Theta\Gamma_{h_n}^{-1}x, \Gamma_{h_n}^{-1}x\rangle$ for $x \in \mathbb{H}$, where $\Theta \equiv \mathsf{E}[\{(X^{\otimes 2}-\Gamma)\beta\}^{\otimes 2}]$. Note that

$$\begin{aligned}\gamma_j^{-1}\|\Lambda^{1/2}\phi_j\|^2 &= \gamma_j^{-1}\langle\Lambda\phi_j,\phi_j\rangle = \gamma_j^{-1}\langle\mathsf{E}[X^{\otimes 2}\varepsilon^2]\phi_j,\phi_j\rangle = \gamma_j^{-1}\mathsf{E}[\langle X,\phi_j\rangle^2\varepsilon^2] \\ &= \mathsf{E}[\xi_j^2\sigma^2(X)] = \sum_{l=1}^\infty \gamma_l\rho_l^2\mathsf{E}[\xi_j^2\xi_l^2] \le (\sup_{j\in\mathbb{N}}\mathsf{E}[\xi_j^4])\sum_{l=1}^\infty\gamma_l\rho_l^2 < \infty\end{aligned}$$

for each $j \in \mathbb{N}$. Thus, the second condition in Proposition 30 is satisfied. Recall from Lemma 46 that in this construction of the FPC scores, $s_{h_n}((I-\Pi_{g_n})X_0) = s_{h_n}(X_0) - s_{g_n}(X_0)$. By the result in Lemma 46, we have

$$\begin{aligned}(h_n-g_n)^{-1}s_{h_n}((I-\Pi_{g_n})X_0) &= \frac{h_n}{h_n-g_n}\frac{s_{h_n}(X_0)}{h_n} - \frac{g_n}{h_n-g_n}\frac{s_{g_n}(X_0)}{g_n} \\ &\xrightarrow{\mathsf{P}} \left(\frac{1}{1-\tau^{-1}} - \frac{1}{\tau-1}\right)\mathsf{E}[\xi^4]\left(\sum_{j=1}^\infty\gamma_j\rho_j^2\right)\xi^2 = \mathsf{E}[\xi^4]\left(\sum_{j=1}^\infty\gamma_j\rho_j^2\right)\xi^2.\end{aligned}$$

Similarly, by using the result in Lemma 47, we have
$(h_n-g_n)^{-1}r_{h_n}((I-\Pi_{g_n})X_0) \xrightarrow{\mathsf{P}} \mathsf{E}[\xi^4]\|\Gamma^{1/2}\beta\|^2\xi^2$. This implies that

$$\begin{aligned}&(h_n-g_n)^{-1}\{s_{h_n}((I-\Pi_{g_n})X_0) + r_{h_n}((I-\Pi_{g_n})X_0)\} \\ &\xrightarrow{\mathsf{P}} \mathsf{E}[\xi^4]\left\{\left(\sum_{j=1}^\infty\gamma_j\rho_j^2\right) + \|\Gamma^{1/2}\beta\|^2\right\}\xi^2\end{aligned}$$

and hence this satisfies the third condition in Proposition 30. Also, the last condition in Proposition 30 is guaranteed by Lemma 48. Finally, by Lemma 46 and Lemma 47, the ratio of the scalings converges as

$$\begin{aligned}&\frac{r_{h_n}((I-\Pi_{g_n})X_0) + s_{h_n}((I-\Pi_{g_n})X_0)}{s_{h_n}(X_0)} \\ =&\frac{h_n^{-1}r_{h_n}((I-\Pi_{g_n})X_0) + h_n^{-1}s_{h_n}((I-\Pi_{g_n})X_0)}{h_n^{-1}s_{h_n}(X_0)} \\ &\xrightarrow{\mathsf{P}} \frac{\mathsf{E}[\xi^4]\|\Gamma^{1/2}\beta\|^2(1-\tau^{-1})\xi^2 + \mathsf{E}[\xi^4]\left(\sum_{j=1}^\infty\gamma_j\rho_j^2\right)(1-\tau^{-1})\xi^2}{\mathsf{E}[\xi^4]\left(\sum_{j=1}^\infty\gamma_j\rho_j^2\right)\xi^2} \\ &= (1-\tau^{-1})\left(\|\Gamma^{1/2}\beta\|^2\Big/\sum_{j=1}^\infty\gamma_j\rho_j^2 + 1\right),\end{aligned}$$

and we have the desired result by Slutsky's theorem. □

**Lemma 49.** *Suppose that Condition (A6) holds. For any sequence* $\{\zeta_h\}_{h\in\mathbb{N}}$ *such that* $\{h^{-1}\zeta_h\}_{h\in\mathbb{N}}$ *is non-decreasing, we have the following moment inequality:*

$$\frac{n}{h_n}\mathsf{E}[\langle\Gamma_{h_n}^{-1}(I-\Pi_{g_n})\Gamma\beta, X_0\rangle^2] \le \frac{n}{\zeta_{g_n}}\frac{g_n}{h_n}\left(\sum_{j>g_n}^{h_n}\gamma_j\right)\sup_{j\in\mathbb{N}}\left(j^{-1}\zeta_j\beta_j^2\right).$$

*Hence, if* $\sup_{j\in\mathbb{N}}\left(j^{-1}\zeta_j\beta_j^2\right)<\infty$ *and* $n=O(\zeta_{g_n})$ *as* $n\to\infty$*, then we have that*

$$\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\Gamma_{h_n}^{-1}(I-\Pi_{g_n})\Gamma\beta, X_0\rangle \xrightarrow{\mathsf{P}} 0.$$

*Proof.* Since

$$\begin{aligned}\frac{n}{h_n}\mathsf{E}[\langle\Gamma_{h_n}^{-1}(I-\Pi_{g_n})\Gamma\beta, X_0\rangle^2] &= \frac{n}{h_n}\sum_{j>g_n}^{h_n}\gamma_j\beta_j^2 = \frac{n}{h_n}\sum_{j>g_n}^{h_n}\gamma_j(j^{-1}\zeta_j)^{-1}(j^{-1}\zeta_j)\beta_j^2\\ &\le \frac{n}{\zeta_{g_n}}\frac{g_n}{h_n}\left(\sum_{j>g_n}^{h_n}\gamma_j\right)\sup_{j\in\mathbb{N}}\left(j^{-1}\zeta_j\beta_j^2\right)=O(1)o(1)=o(1)\end{aligned}$$

it follows that

$$\begin{aligned}\frac{n}{s_{h_n}(X_0)}\langle\Gamma_{h_n}^{-1}(I-\Pi_{g_n})\Gamma\beta, X_0\rangle^2 &= \frac{h_n}{s_{h_n}(X_0)}\frac{n}{h_n}\langle\Gamma_{h_n}^{-1}(I-\Pi_{g_n})\Gamma\beta, X_0\rangle^2\\ &= O_{\mathsf{P}}(1)o_{\mathsf{P}}(1)=o_{\mathsf{P}}(1).\end{aligned}$$

□

### 3.10.7 Failure of both paired bootstrap methods

This section treat the failure of both paired bootstrap methods when $\tau\equiv\lim_{n\to\infty}h_n/g_n\in(0,1)$ described in Proposition 14 in the main paper.

*Proof of Proposition 14.* By Lemma 46, with the same argument as the one in the proof of Proposition 31, one can show that the sufficient conditions in Lemma 34: $(g_n-h_n)s_{g_n}((I-\Pi_{h_n})X_0)^{-1}=O_{\mathsf{P}}(1)$ and

$$\frac{(g_n-h_n)^{-1}s_{g_n}((I-\Pi_{h_n})X_0)}{h_n^{-1}s_{h_n}(X_0)}\xrightarrow{\mathsf{P}} 1.$$

Based on the bootstrap theory developed in Sections 3.10.3-3.10.4, the (modified) bootstrap statistics $T_n^*(X_0)$ is decomposed as $T_n^*(X_0) = T_{\text{var},n}^* + B_{\text{dim},n}$, where $\sup_{y\in\mathbb{R}} |\mathsf{P}^*(T_{\text{var},n}^* \le y|X_0) - \Phi(y)| \xrightarrow{\mathsf{P}} 0$ and $\sup_{y\in\mathbb{R}} |\mathsf{P}(B_{\text{dim},n} \le y|X_0) - \Phi(y/\sigma_{\text{dim}}(\tau))| \xrightarrow{\mathsf{P}} 0$ with $\sigma_{\text{dim}}^2(\tau) \equiv \tau^{-1} - 1$. By using the same argument as the one in the proof of Proposition 13 in the main paper, we derive

$$\mathsf{P}^*(T_n^*(X_0) \le y|X_0) - \mathsf{P}(T_n(X_0) \le y|X_0) \xrightarrow{\mathsf{d}} \Phi(y + \sigma_{\text{dim}}(\tau)Z) - \Phi(y), \quad y \in \mathbb{R},$$

as elements in $\mathbb{D}$. This result also holds for the naive bootstrap construction $T_{n,naive}^*(X_0)$, since it is equal to $T_n^*(X_0)$ when $h_n < g_n$. This completes the proof. □

## 3.11 Additional simulation results

This section provides further simulation results in addition to those in Section 3.5 of the main paper and the detailed simulation procedures. Section 3.11.1 contains additional results of empirical coverage simulation under extra scenarios while further empirical rejection rates are given in Section 3.11.2. Meanwhile, in Section 3.11.3, we provide further results for the failure of both modified and naive paired bootstrap methods.

### 3.11.1 Empirical coverage probabilities of bootstrap intervals

In addition to set-ups for simulation in Section 3.5.1 of the main paper, we consider different choices of distribution of the latent variable $\xi$ among $t(4)$, $t(5)$, $t(7)$, $t(9)$, and $\mathsf{N}(0,1)$. Different decay rates $a, b \in \{1.5, 2.5, 3.5, 4.5, 5.5\}$ for the regressor and the slope function respectively are considered. Another error distribution is considered in addition to the centered chi-square distribution described in Section 3.5.1 of the main paper: for a given regressor $X$, the dependent error $\varepsilon$ follows the centered uniform distribution $\mathsf{U}(-a(X), a(X))$ where $a(X) = \sqrt{3}\langle X, \rho\rangle$ with a fixed function $\rho \in L^2([0,1])$. Here, $\rho(t) = t^3 - 1.5t - 2.5$ is used. In this case, the marginal variance is $\mathsf{var}[\varepsilon] = \langle \Gamma\rho, \rho\rangle = \sum_{j=1}^{\infty} \gamma_j \langle \rho, \phi_j\rangle^2$, and hence, the we generate the corresponding independent error from the centered uniform distribution $\mathsf{U}(-a, a)$ with $a = \sqrt{3\langle \Gamma\rho, \rho\rangle}$. Due to similarity and brevity, we report only partial results when $a \in \{2.5, 3.5, 4.5\}$, $b = 5.5$, and the errors follow the centered chi-square distribution as described in Section 3.5.1 of the main paper.

We provide the details of simulation algorithm to examine the empirical coverage probabilities and average widths for intervals, as these are not included in Section 3.5.1 of the main paper. At each Monte Carlo iteration, we simulate the new predictor $X_0$ as well as the data samples $\{(X_i, Y_i)\}_{i=1}^n$. Here, the Monte Carlo simulation size $M$ and the bootstrap resample size $Q$ are given as $M = 1000 = Q$. The intervals are obtained from either residual or paired bootstrap method. We refer to [53] and its supplement for the description of the estimates used in the residual bootstrap method.

The simulation is conducted as follows. For each $m = 1, \ldots, M$, perform the following.

1. (Simulation) Simulate $X_i$ and $\varepsilon_i$ with $X_i \overset{\text{d}}{=} X$ and $\varepsilon_i \overset{\text{d}}{=} \varepsilon$ respectively, where the pairs $\{(X_i, \varepsilon_i)\}_{i=1}^n$ are independent but $\varepsilon$ may be dependent of $X$, and compute the response $Y_i = \langle \beta, X_i \rangle + \varepsilon_i$ for $i = 1, \ldots, n$.

2. (Residual bootstrap)

   (a) (Residuals) Compute the residuals $\hat{\varepsilon}_{i,k_n} = Y_i - \langle \hat{\beta}_{k_n}, X_i \rangle$ for $i = 1, \ldots, n$.

   (b) (Residual bootstrap) To approximate the residual bootstrap distribution, do the following for $q = 1, \ldots, Q$.

      i. Draw independent bootstrap errors $\{\varepsilon_{q,i}^*\}_{i=1}^n$ from the uniform distribution on the centered residuals $\{\hat{\varepsilon}_{i,k_n} - \bar{\hat{\varepsilon}}_{k_n}\}_{i=1}^n$.

      ii. Compute the bootstrap responses $Y_{q,i}^* = \langle \hat{\beta}_{g_n}, X_i \rangle + \varepsilon_{q,i}^*$, and construct the bootstrap estimate $\hat{\beta}_{q,h_n}^*$ based on the bootstrap samples $\{(X_i, Y_{q,i}^*)\}_{i=1}^n$.

      iii. Compute the bootstrap statistic

$$\hat{T}_{q,n}^{RB*}(X_0) \equiv \sqrt{\frac{n}{\hat{t}_{h_n}(X_0)}}[\langle \hat{\beta}_{q,h_n}^*, X_0 \rangle - \langle \hat{\beta}_{g_n}, X_0 \rangle],$$

      where the scaling $\hat{t}_{h_n}$ is defined in Equation (8) of [53] as $\hat{t}_{h_n}(x) = \sum_{j=1}^{h_n} \hat{\gamma}_j^{-1} \langle x, \hat{\phi}_j \rangle^2$ for $x \in \mathbb{H}$.

3. (Paired bootstrap)

   (a) (Bias correction term) Compute the bias correction term $\hat{U}_{n,g_n} = n^{-1}\sum_{i=1}^n (X_i - \bar{X})(\hat{\varepsilon}_{i,g_n} - \bar{\hat{\varepsilon}}_{g_n})$, where $\hat{\varepsilon}_{i,g_n} = Y_i - \langle \hat{\beta}_{g_n}, X_i \rangle$ are the residuals for $i = 1, \ldots, n$ with its average $\bar{\hat{\varepsilon}}_{g_n} = n^{-1}\sum_{i=1}^n \hat{\varepsilon}_{i,g_n}$.

   (b) (Proposed paired bootstrap) To approximate the paired bootstrap distribution, do the following for $q = 1, \ldots, Q$.

      i. Draw independent bootstrap pairs $\{(X_i^*, Y_i^*)\}_{i=1}^n$ from the uniform distribution on the samples $\{(X_i, Y_i)\}_{i=1}^n$.

      ii. Compute the bias corrected bootstrap estimate $\hat{\beta}^*_{q,h_n}$ and the bootstrap scaling $\hat{s}^*_{q,h_n}(X_0)$ based on the bootstrap samples $\{(X_i^*, Y_i^*)\}_{i=1}^n$ and $\hat{\beta}^*_{q,h_n}$.

      iii. Compute the bootstrap statistics either with or without studentization:

$$\hat{T}^{PB*}_{q,n}(X_0) \equiv \sqrt{\frac{n}{\hat{s}_{h_n}(X_0)}}[\langle \hat{\beta}^*_{q,h_n}, X_0 \rangle - \langle \hat{\beta}_{g_n}, X_0 \rangle],$$

$$\hat{T}^{PB_{std}*}_{q,n}(X_0) \equiv \sqrt{\frac{n}{\hat{s}^*_{q,h_n}(X_0)}}[\langle \hat{\beta}^*_{q,h_n}, X_0 \rangle - \langle \hat{\beta}_{g_n}, X_0 \rangle].$$

   (c) (Naive paired bootstrap) To approximate the paired bootstrap distribution, do the following for $q = 1, \ldots, Q$.

      i. Draw independent bootstrap pairs $\{(X_i^*, Y_i^*)\}_{i=1}^n$ from the uniform distribution on the samples $\{(X_i, Y_i)\}_{i=1}^n$.

      ii. Compute the bootstrap estimate $\hat{\beta}^*_{q,h_n,naive}$ without bias correction and the bootstrap scaling $\hat{s}^*_{q,h_n,naive}(X_0)$ based on the bootstrap samples $\{(X_i^*, Y_i^*)\}_{i=1}^n$ and $\hat{\beta}^*_{q,h_n,naive}$.

      iii. Compute the bootstrap statistics either with or without studentization:

$$\hat{T}^{PB*}_{q,n,naive}(X_0) \equiv \sqrt{\frac{n}{\hat{s}_{h_n}(X_0)}}[\langle \hat{\beta}^*_{q,h_n,naive}, X_0 \rangle - \langle \hat{\beta}_{g_n}, X_0 \rangle],$$

$$\hat{T}^{PB_{std}*}_{q,n,naive}(X_0) \equiv \sqrt{\frac{n}{\hat{s}^*_{q,h_n,naive}(X_0)}}[\langle \hat{\beta}^*_{q,h_n,naive}, X_0 \rangle - \langle \hat{\beta}_{g_n}, X_0 \rangle].$$

4. For all cases, construct the following confidence intervals for $\langle \beta, X_0 \rangle$

(a) (CLT)

The (symmetrized) confidence interval for $\langle \beta, X_0 \rangle$ from CLT is

$$CI_{CLT} = \left[ \langle \hat{\beta}_{h_n}, X_0 \rangle - \sqrt{\frac{\hat{t}_{h_n}(X_0)}{n}} z_{1-\alpha/2}, \langle \hat{\beta}_{h_n}, X_0 \rangle + \sqrt{\frac{\hat{t}_{h_n}(X_0)}{n}} z_{1-\alpha/2} \right],$$

where $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution.

(b) (Residual bootstrap)

i. (Unsymmetrized intervals) Compute the $\alpha/2$ and $1-\alpha/2$ quantiles of $\{\hat{T}_{q,n}^{RB*}(X_0)\}_{q=1}^{Q}$, say $l$ and $u$. Then, the unsymmetrized confidence interval for $\langle \beta, X_0 \rangle$ is

$$CI_{RB,unsym} = \left[ \langle \hat{\beta}_{h_n}, X_0 \rangle - \sqrt{\frac{\hat{t}_{h_n}(X_0)}{n}} u, \langle \hat{\beta}_{h_n}, X_0 \rangle - \sqrt{\frac{\hat{t}_{h_n}(X_0)}{n}} l \right]$$

ii. (Symmetrized intervals) Compute the $1-\alpha/2$ quantile of $\{|\hat{T}_{q,n}^{RB*}(X_0)|\}_{q=1}^{Q}$, say $u$. Then, the symmetrized confidence interval for $\langle \beta, X_0 \rangle$ is

$$CI_{RB,sym} = \left[ \langle \hat{\beta}_{h_n}, X_0 \rangle - \sqrt{\frac{\hat{t}_{h_n}(X_0)}{n}} u, \langle \hat{\beta}_{h_n}, X_0 \rangle + \sqrt{\frac{\hat{t}_{h_n}(X_0)}{n}} u \right]$$

(c) (Proposed paired bootstrap without studentization)

Replace $\hat{T}_{q,n}^{RB*}(X_0)$ and $\hat{t}_{h_n}(X_0)$ by $\hat{T}_{q,n}^{PB*}(X_0)$ and $\hat{s}_{h_n}(X_0)$ respectively in the procedure (b) to obtain $CI_{PB,unsym}$ and $CI_{PB,sym}$.

(d) (Proposed paired bootstrap with studentization)

Replace $\hat{T}_{q,n}^{PB*}(X_0)$ by $\hat{T}_{q,n}^{PB_{std}*}(X_0)$ in the procedure (c) to obtain $CI_{PB_{std},unsym}$ and $CI_{PB_{std},sym}$.

(e) (Naive paired bootstrap either with or without studentization)

In the above procedures in (c) and (d), replace $\hat{T}_{q,n}^{PB*}(X_0)$ and $\hat{T}_{q,n}^{PB_{std}*}(X_0)$ by $\hat{T}_{q,n,naive}^{PB*}(X_0)$ and $\hat{T}_{q,n,naive}^{PB_{std}*}(X_0)$ to obtain $CI_{naivePB,unsym}$, $CI_{naivePB,sym}$, $CI_{naivePB_{std},unsym}$, and $CI_{naivePB_{std},sym}$.

5. Let $CI$ denote one of the intervals constructed above. Compute $I_m = \mathbb{I}(\langle \beta, X_0 \rangle \in CI)$.

The coverage probability $1-\alpha$ is then approximated by $M^{-1}\sum_{m=1}^{M} I_m$.

Figures 3.7-3.16 provide the empirical coverage probabilities and average widths of each interval with different tuning parameter $h_n$ under the scenarios considered.

### 3.11.2 Empirical rejection rates of bootstrap testing

We describe the exact procedure of the simulation to obtain empirical rejection rates of the bootstrap testing and provide additional results to those in Section 3.5.2. In simulation, we generate the new predictors $\{X_{0,l}^p\}_{l=1}^L$ and fix them before the Monte Carlo iteration. The Monte Carlo simulation size $M$ and the bootstrap sample size $Q$ are again provided as $M = 1000 = Q$. For each $m = 1, \ldots, M$, perform the following.

1. (Simulation) Simulate $X_i$ and $\varepsilon_i$ with $X_i \overset{\mathrm{d}}{=} X$ and $\varepsilon_i \overset{\mathrm{d}}{=} \varepsilon$ respectively, where the pairs $\{(X_i, \varepsilon_i)\}_{i=1}^n$ are independent but $\varepsilon$ may be dependent of $X$, and compute the response $Y_i = \langle \beta, X_i \rangle + \varepsilon_i$ for $i = 1, \ldots, n$.

2. (Test statistics) Compute the $L^2$ and maximum type test statistics

$$W_{n,L^2} \equiv \sum_{l=1}^{L} \left[\hat{T}_{n,l}^{H_0}\right]^2 \quad \text{and} \quad W_{n,\max} \equiv \max_{1\le l\le L} \left|\hat{T}_{n,l}^{H_0}\right|,$$

   where $\hat{T}_{n,l}^{H_0} \equiv \sqrt{n/\hat{s}_{h_n}(X_{0,l})}\langle \hat{\beta}_{h_n}, X_{0,l}\rangle$ for $l = 1, \ldots, L$.

3. (Paired bootstrap when not enforcing the null)

   (a) (Bias correction term) Compute the bias correction term $\hat{U}_{n,g_n} = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})(\hat{\varepsilon}_{i,g_n} - \bar{\hat{\varepsilon}}_{g_n})$, where $\hat{\varepsilon}_{i,g_n} = Y_i - \langle \hat{\beta}_{g_n}, X_i\rangle$ are the residuals for $i = 1, \ldots, n$ with its average $\bar{\hat{\varepsilon}}_{g_n} = n^{-1}\sum_{i=1}^{n} \hat{\varepsilon}_{i,g_n}$.

   (b) (Paired bootstrap when not enforcing the null) To approximate the paired bootstrap distribution, do the following for $q = 1, \ldots, Q$.

      i. Draw independent bootstrap paris $\{(X_i^*, Y_i^*)\}_{i=1}^n$ from the uniform distribution on the samples $\{(X_i, Y_i)\}_{i=1}^n$.

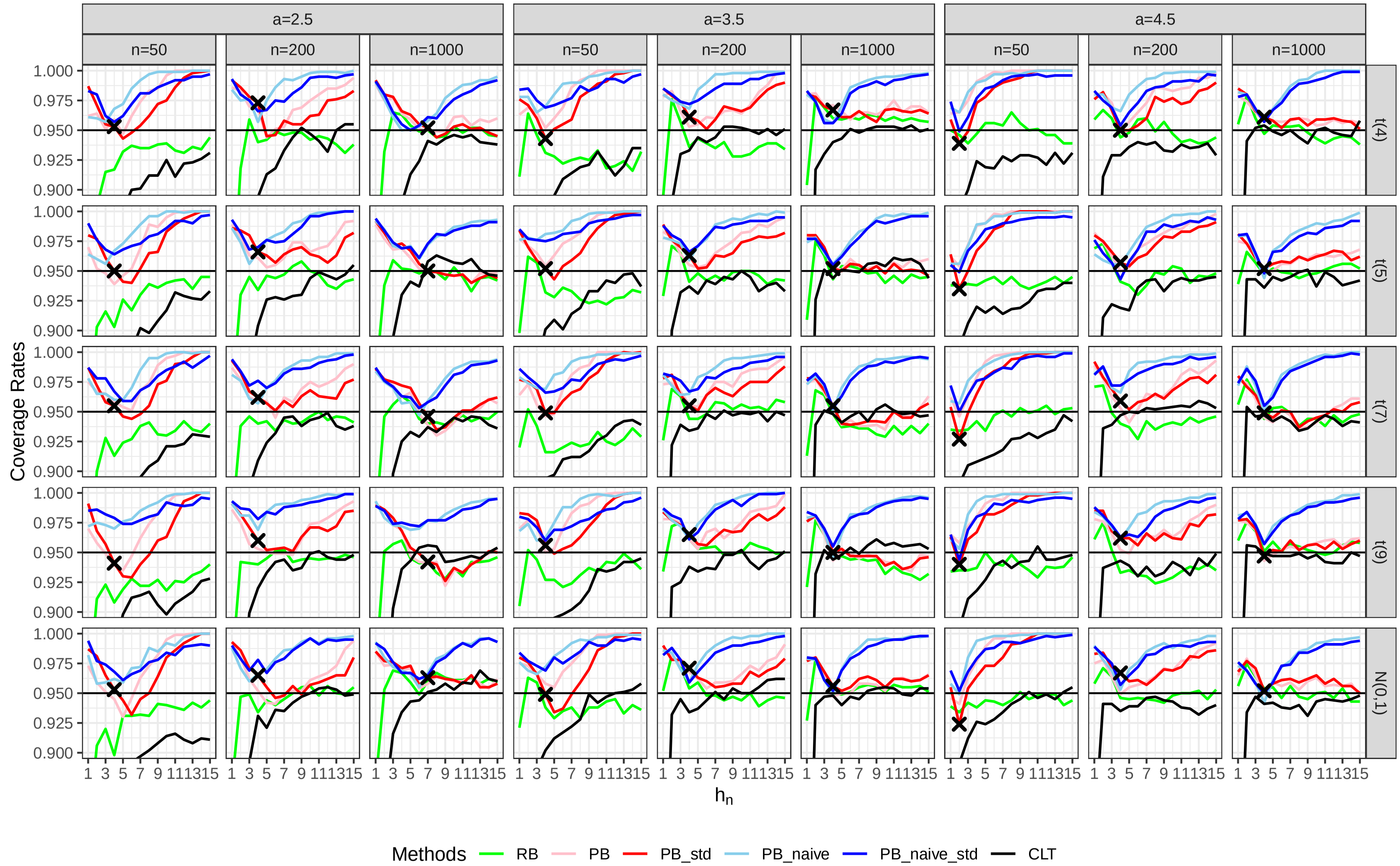


Figure 3.7: Empirical coverage rates of symmetrized confidence intervals for $\langle \beta, X_0 \rangle$ from each method over different truncation levels $h_n$ and $g_n = k_n$ when the new predictor $X_0$ is random under the scenarios when the error follows the chi-square distribution and is independent of the regressor. Crosses $\times$ indicate bootstrap coverages with $h_n$ selected by a proposed rule.

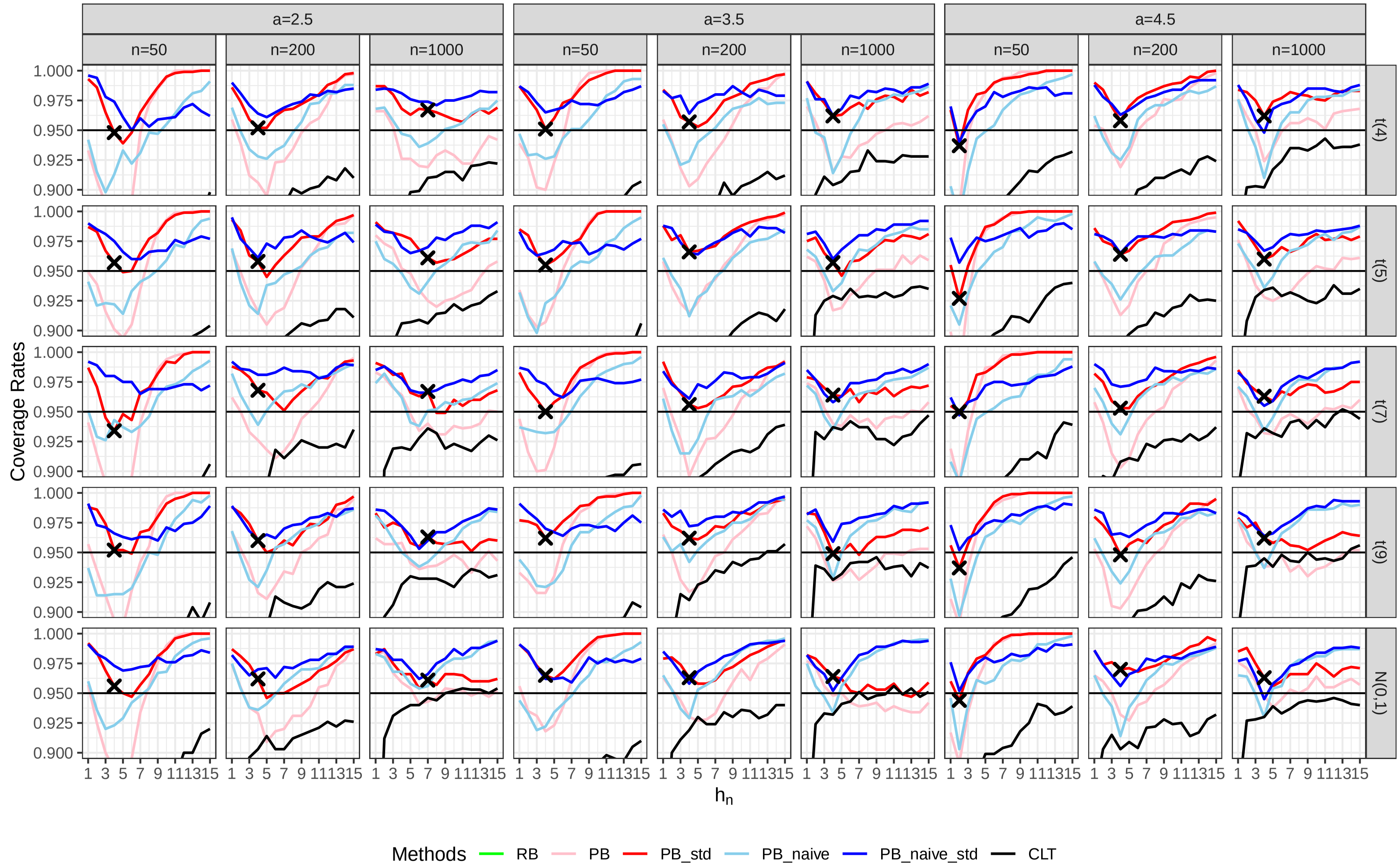


Figure 3.8: Empirical coverage rates of symmetrized confidence intervals for $\langle \beta, X_0 \rangle$ from each method over different truncation levels $h_n$ and $g_n = k_n$ when the new predictor $X_0$ is random under the scenarios when the error follows the chi-square distribution and is not independent of the regressor. Crosses $\times$ indicate bootstrap coverages with $h_n$ selected by a proposed rule.

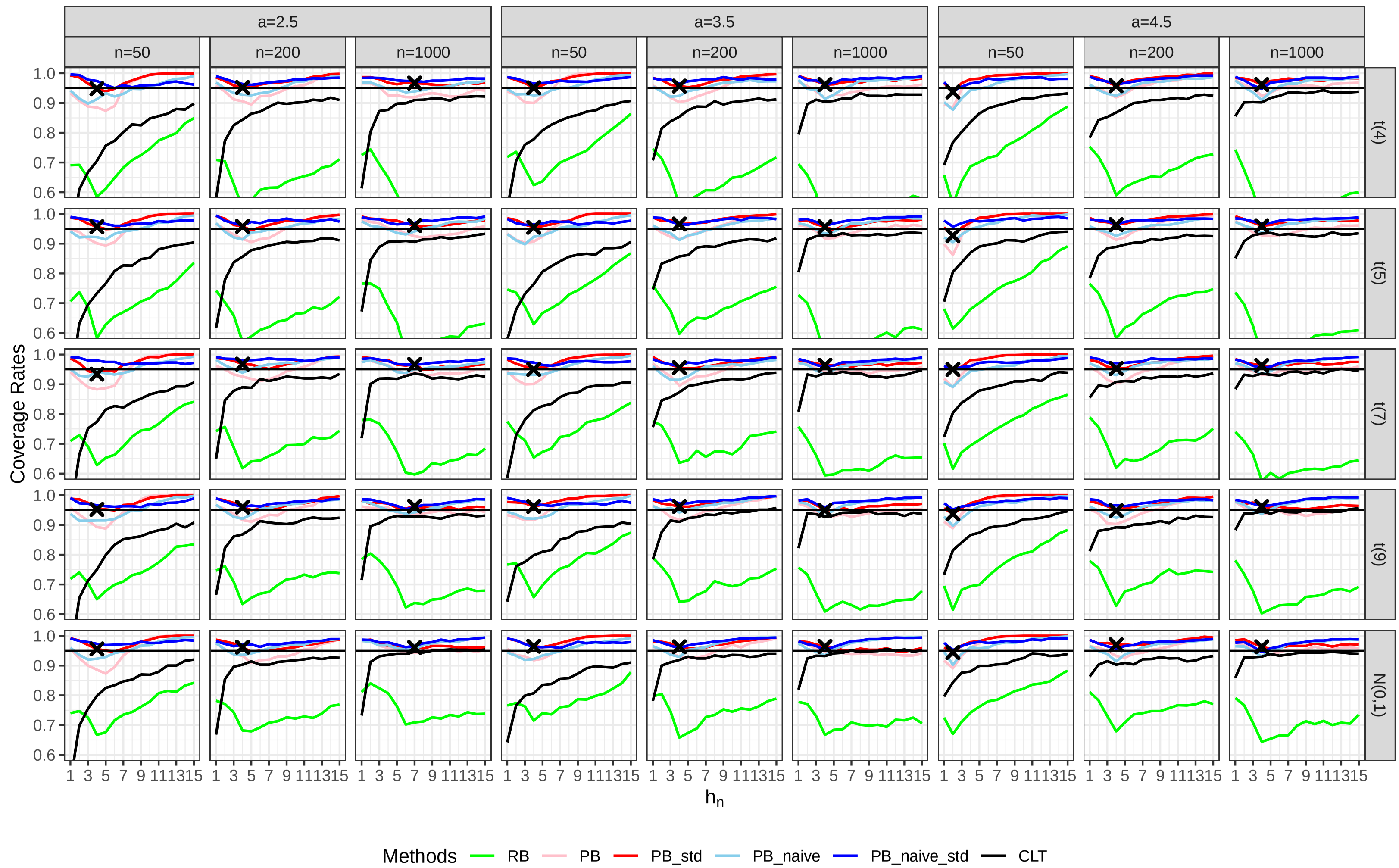


Figure 3.9: A blown-up version of plots in Figure 3.9

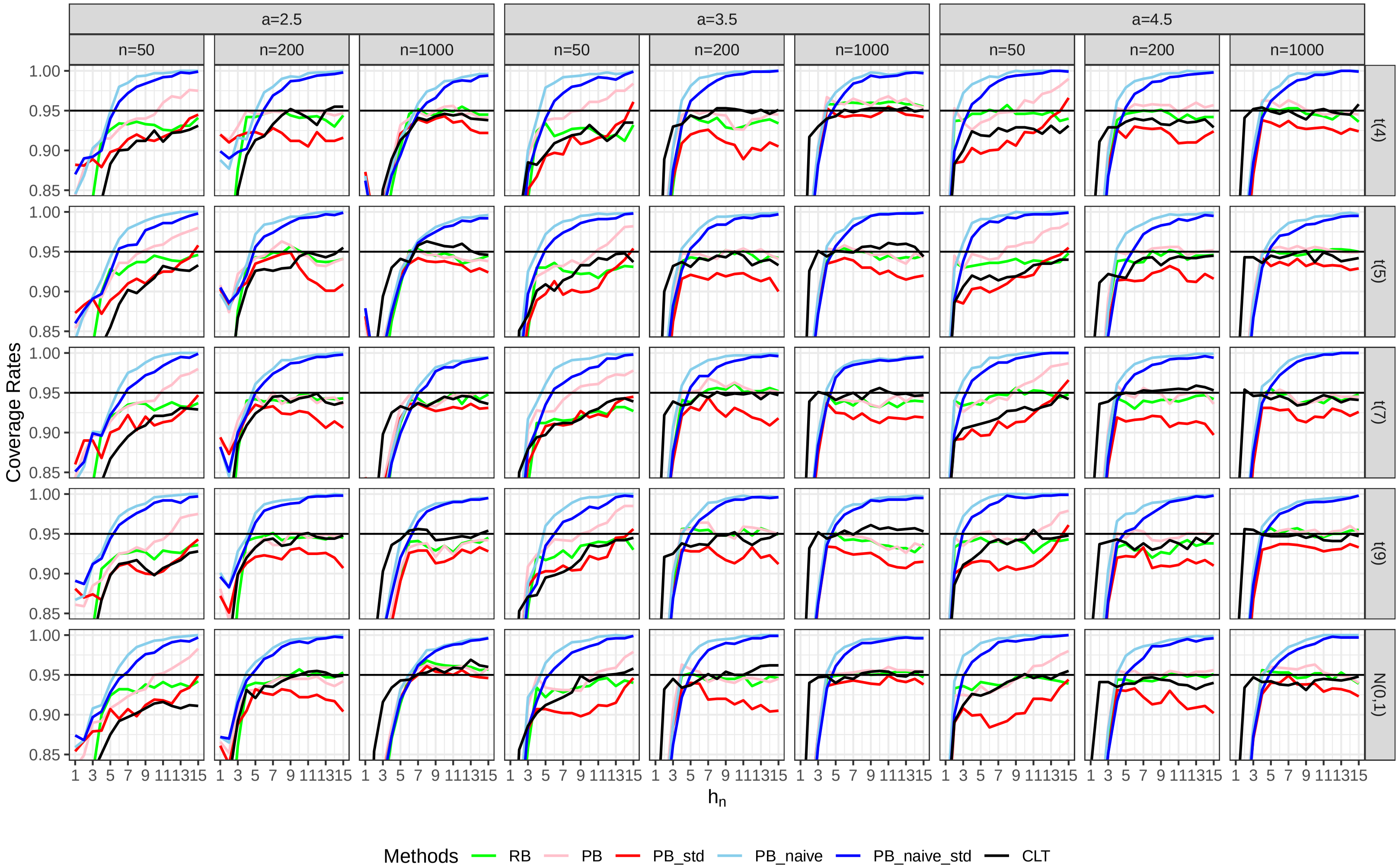


Figure 3.10: Empirical coverage rates of unsymmetrized confidence intervals for $\langle \beta, X_0 \rangle$ from each method over different truncation levels $h_n$ and $g_n = k_n$ when the new predictor $X_0$ is random under the scenarios when the error follows the chi-square distribution and is independent of the regressor.

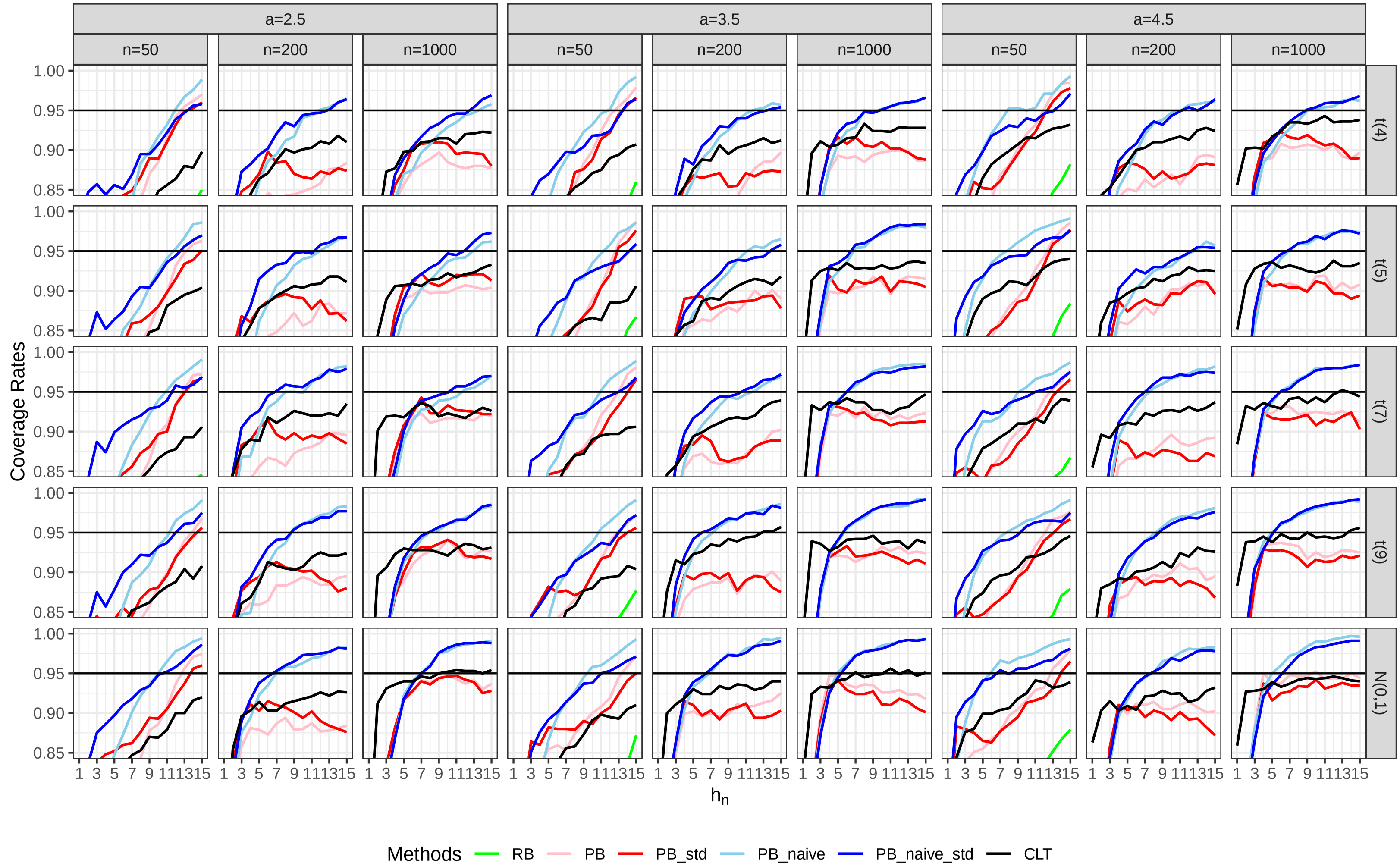


Figure 3.11: Empirical coverage rates of unsymmetrized confidence intervals for $\langle \beta, X_0 \rangle$ from each method over different truncation levels $h_n$ and $g_n = k_n$ when the new predictor $X_0$ is random under the scenarios when the error follows the chi-square distribution and is not independent of the regressor.

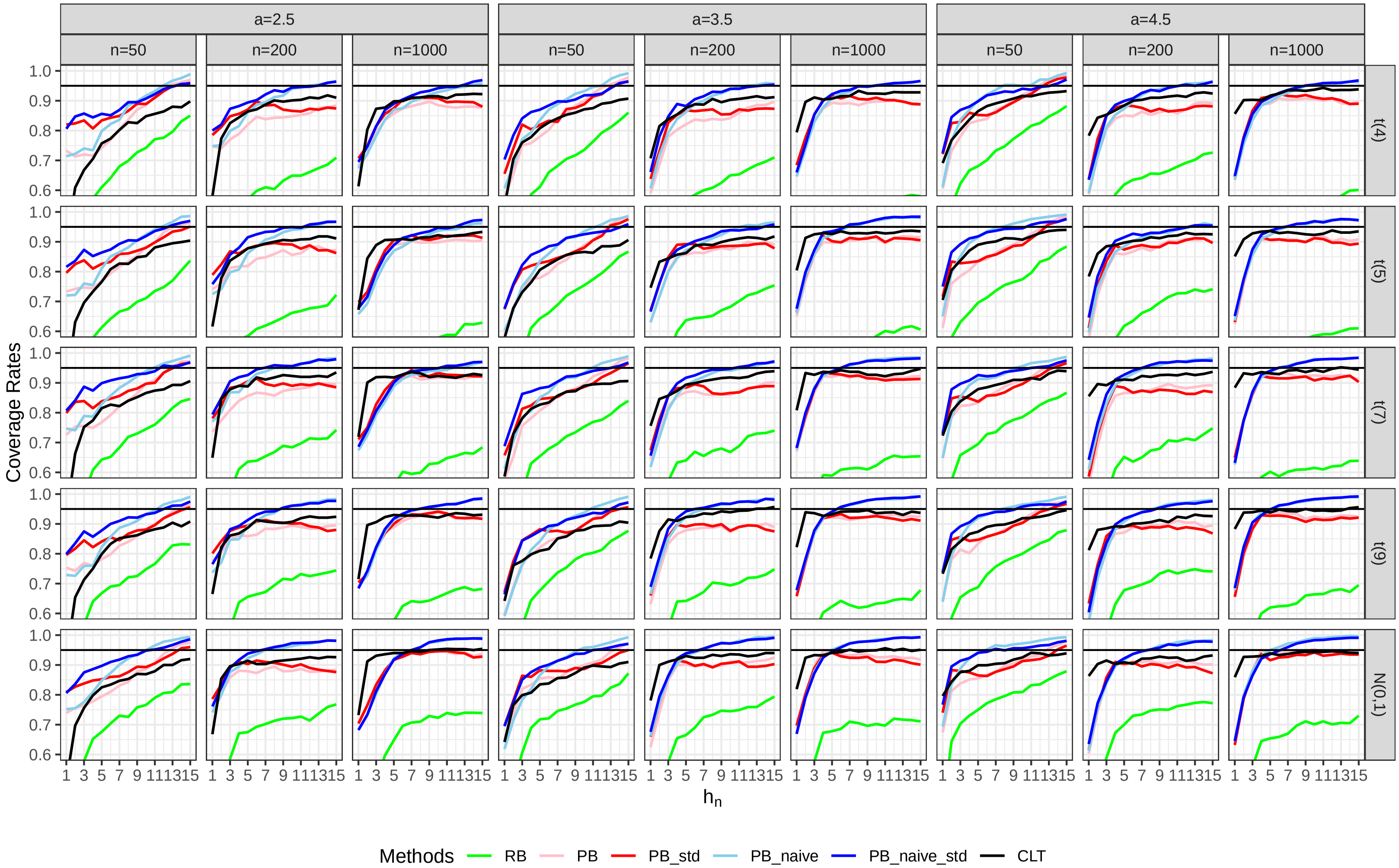


Figure 3.12: A blown-up version of plots in Figure 3.11.

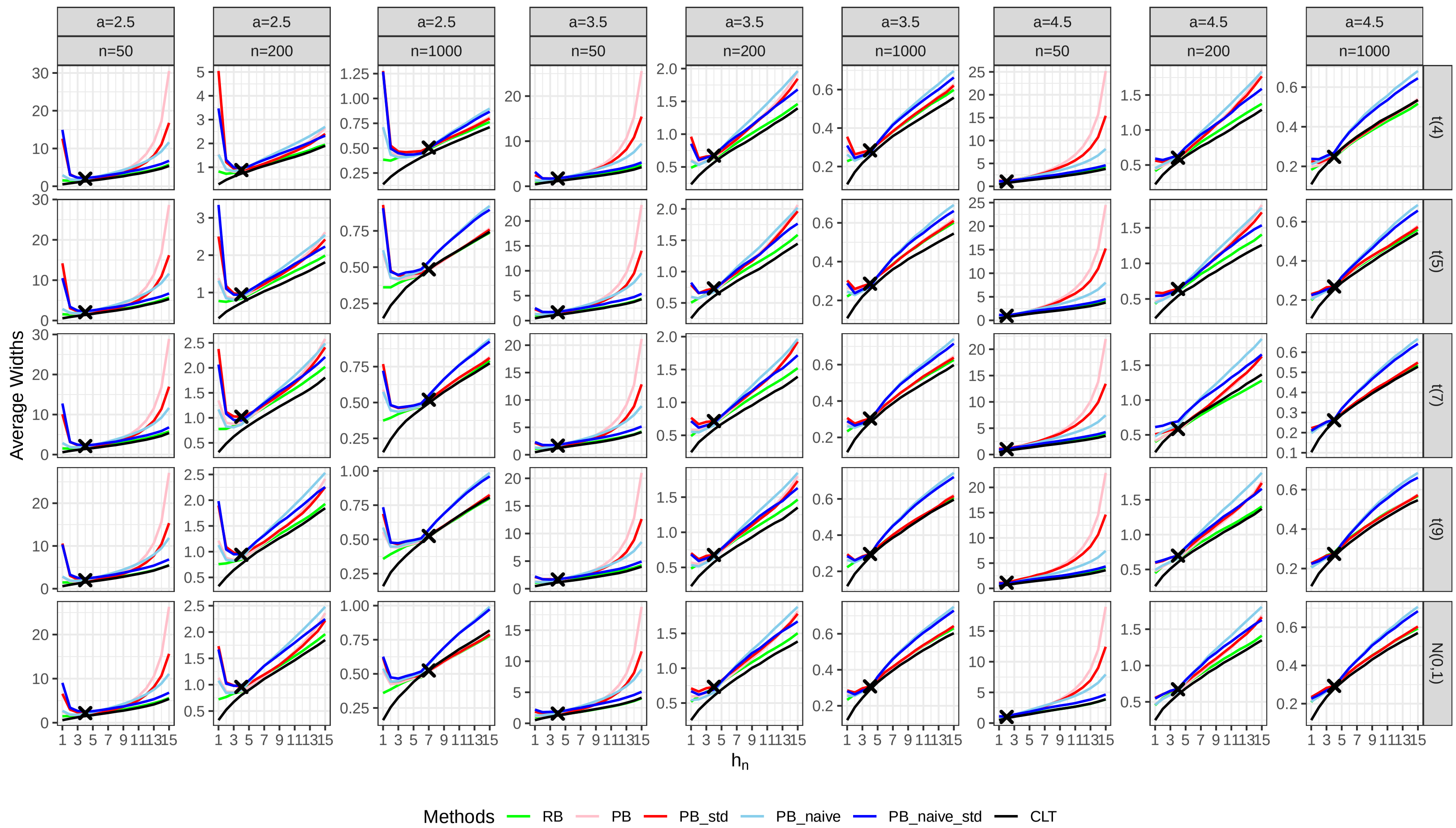


Figure 3.13: Average widths of symmetrized confidence intervals for $\langle \beta, X_0 \rangle$ from each method over different truncation levels $h_n$ and $g_n = k_n$ when the new predictor $X_0$ is random under the scenarios when the error follows the chi-square distribution and is independent of the regressor. Crosses $\times$ indicate bootstrap coverages with $h_n$ selected by a proposed rule.

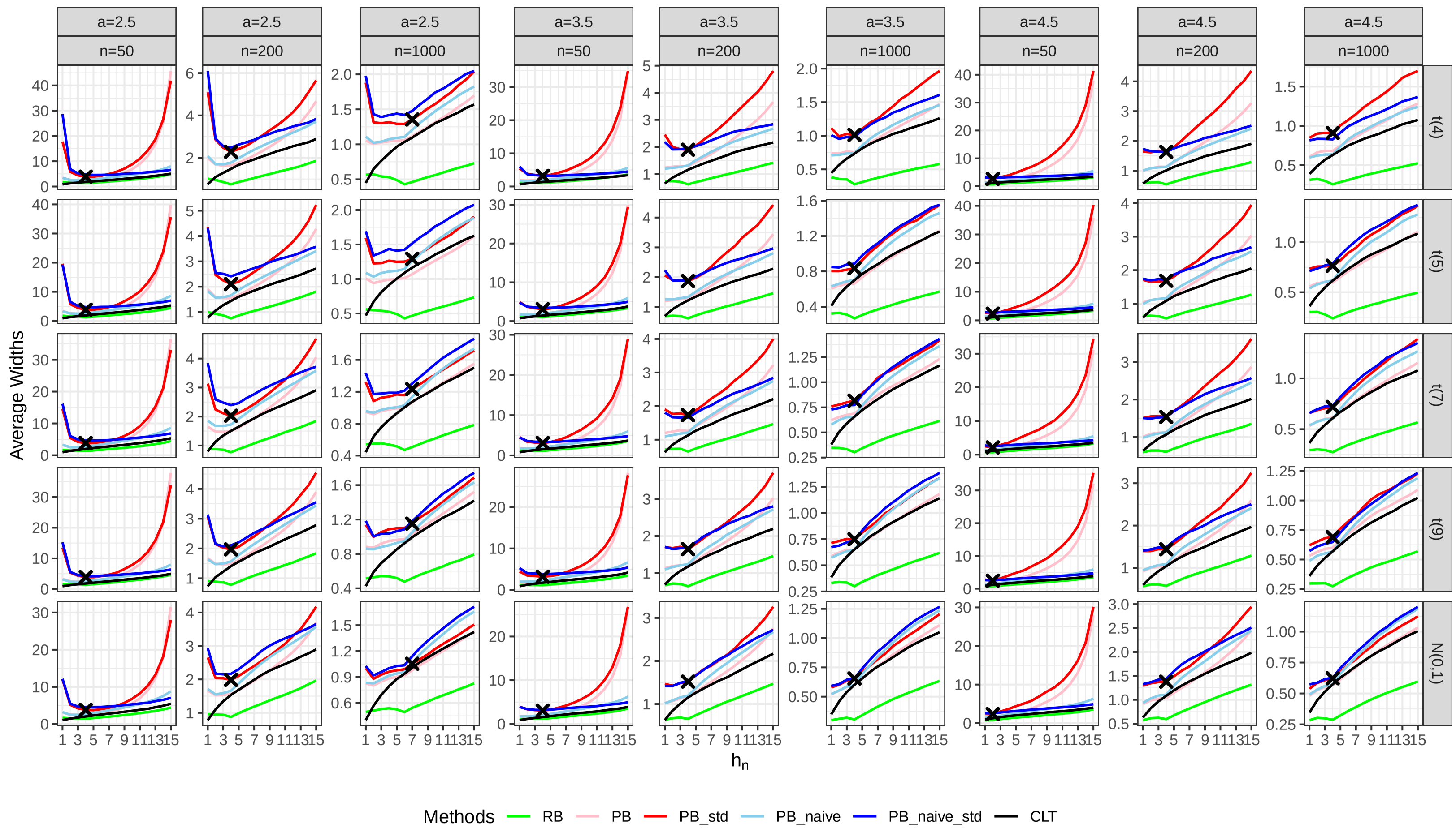


Figure 3.14: Average widths of symmetrized confidence intervals for $\langle \beta, X_0 \rangle$ from each method over different truncation levels $h_n$ and $g_n = k_n$ when the new predictor $X_0$ is random under the scenarios when the error follows the chi-square distribution and is not independent of the regressor. Crosses $\times$ indicate bootstrap coverages with $h_n$ selected by a proposed rule.

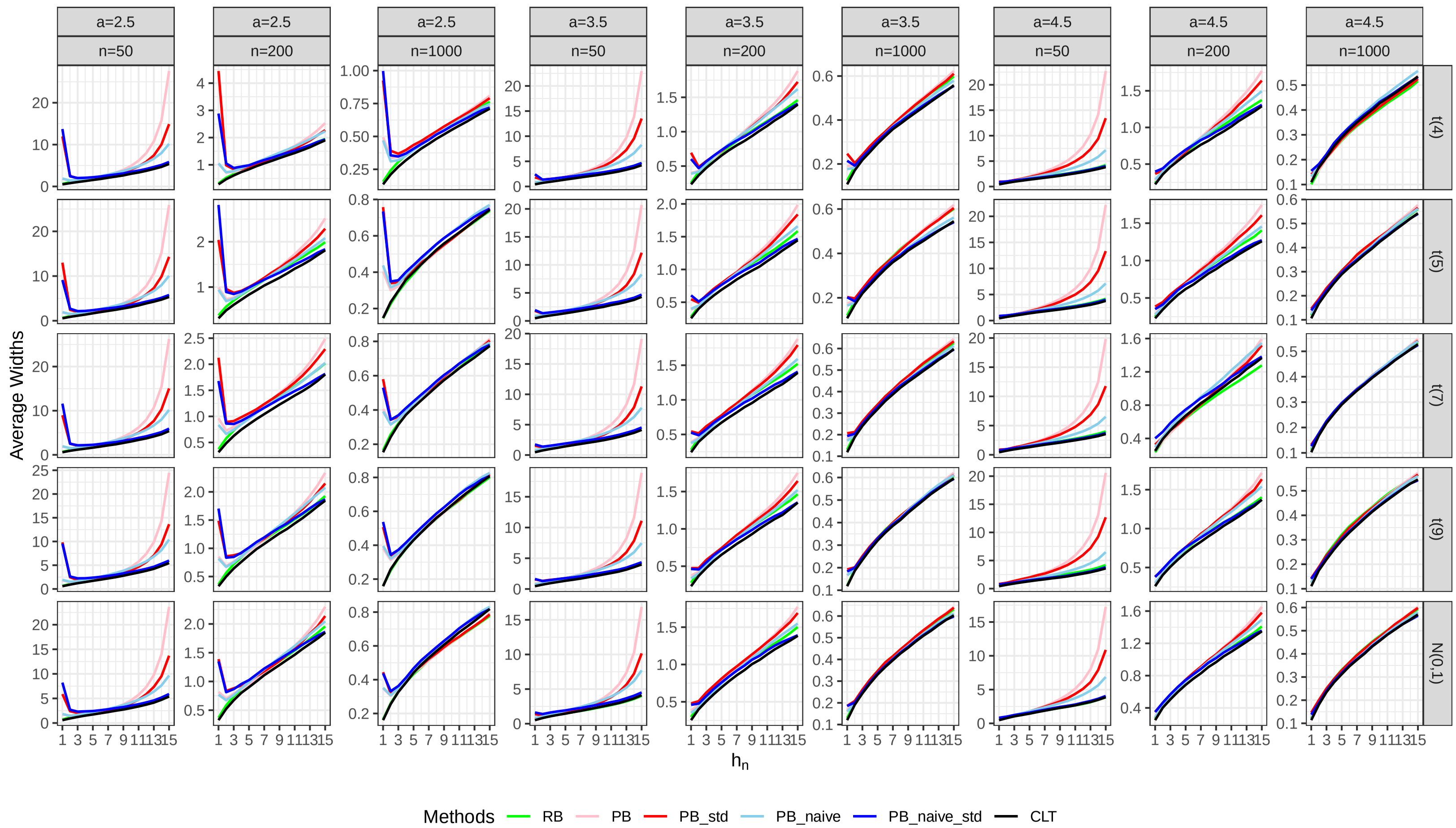


Figure 3.15: Average widths of unsymmetrized confidence intervals for $\langle \beta, X_0 \rangle$ from each method over different truncation levels $h_n$ and $g_n = k_n$ when the new predictor $X_0$ is random under the scenarios when the error follows the chi-square distribution and is independent of the regressor.

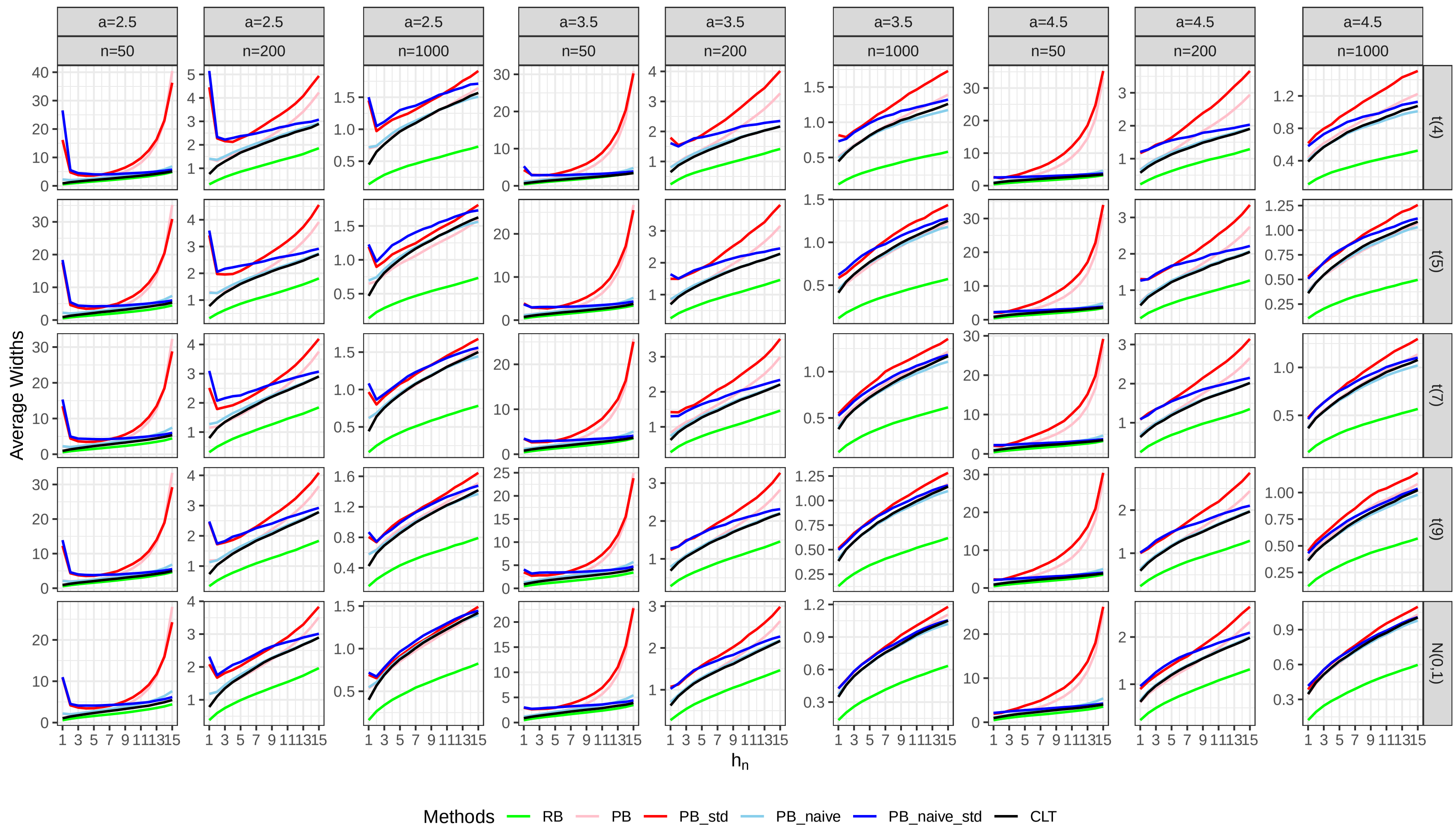


Figure 3.16: Average widths of unsymmetrized confidence intervals for $\langle \beta, X_0 \rangle$ from each method over different truncation levels $h_n$ and $g_n = k_n$ when the new predictor $X_0$ is random under the scenarios when the error follows the chi-square distribution and is not independent of the regressor.

ii. Compute the bias corrected bootstrap estimate $\hat{\beta}^*_{q,h_n}$ and the bootstrap scaling $\hat{s}^*_{q,h_n}(X^p_{0,l})$ based on the bootstrap samples $\{(X^*_i, Y^*_i)\}^n_{i=1}$

iii. Compute the bootstrap statistics with studentization:

$$T^*_{q,n,l} \equiv \sqrt{\frac{n}{\hat{s}^*_{q,h_n}(X^p_{0,l})}}[\langle \hat{\beta}^*_{q,h_n}, X^p_{0,l}\rangle - \langle \hat{\beta}_{g_n}, X^p_{0,l}\rangle]$$

for each $l = 1, \ldots, L$.

iv. Compute the bootstrap $L^2$ and maximum type test statistics

$$W^*_{q,n,L^2} \equiv \sum_{l=1}^{L} \left[\hat{T}^*_{q,n,l}\right]^2 \quad \text{and} \quad W^*_{q,n,\max} \equiv \max_{1\le l\le L} \left|\hat{S}^*_{q,n,l}\right|.$$

4. (Paired bootstrap when enforcing the null)

(a) (Enforcing the null) Compute the estimate $\tilde{\beta}_{g_n} = \hat{\beta}_{g_n} - \Pi_{\mathcal{X}_0}\hat{\beta}_{g_n}$ and the responses $\tilde{Y}_i = Y_i - \langle \Pi_{\mathcal{X}_0}\hat{\beta}_{g_n}, X_i\rangle$ when enforcing the null.

(b) (Bias correction term) Compute the bias correction term $\tilde{U}_{n,g_n} = n^{-1}\sum_{i=1}^n (X_i - \bar{X})(\tilde{\varepsilon}_{i,g_n} - \bar{\tilde{\varepsilon}}_{g_n})$, where $\tilde{\varepsilon}_{i,g_n} = Y_i - \langle \tilde{\beta}_{g_n}, X_i\rangle$ are the residuals for $i = 1, \ldots, n$ with its average $\bar{\tilde{\varepsilon}}_{g_n} = n^{-1}\sum_{i=1}^n \tilde{\varepsilon}_{i,g_n}$.

(c) (Paired bootstrap) To approximate the paired bootstrap distribution, do the following for $q = 1, \ldots, Q$.

i. Draw independent bootstrap paris $\{(\tilde{X}^*_i, \tilde{Y}^*_i)\}^n_{i=1}$ from the uniform distribution on the samples $\{(X_i, \tilde{Y}_i)\}^n_{i=1}$.

ii. Compute the bias corrected bootstrap estimate $\tilde{\beta}^*_{q,h_n}$ and the bootstrap scaling $\tilde{s}^*_{q,h_n}(X^p_{0,l})$ based on the bootstrap samples $\{(\tilde{X}^*_i, \tilde{Y}^*_i)\}^n_{i=1}$

iii. Compute the bootstrap statistics with studentization:

$$T^{*H_0}_{q,n,l} \equiv \sqrt{\frac{n}{\tilde{s}^*_{q,h_n}(X^p_{0,l})}}[\langle \tilde{\beta}^*_{q,h_n}, X^p_{0,l}\rangle - \langle \hat{\beta}_{g_n}, X^p_{0,l}\rangle]$$

for each $l = 1, \ldots, L$.

iv. Compute the bootstrap $L^2$ and maximum type test statistics

$$W^{*H_0}_{q,n,L^2} \equiv \sum_{l=1}^{L} \left[T^{*H_0}_{q,n,l}\right]^2 \quad \text{and} \quad W^{*H_0}_{q,n,\max} \equiv \max_{1\le l\le L} \left|T^{*H_0}_{q,n,l}\right|.$$

5. Let $u$ denote the $1-\alpha$ quantiles of either $\{W^*_{q,n,L^2}\}_{q=1}^Q$, $\{W^*_{q,n,\max}\}_{q=1}^Q$, $\{W^{*H_0}_{q,n,L^2}\}_{q=1}^Q$, or $\{W^{*H_0}_{q,n,\max}\}_{q=1}^Q$, and $W_n$ denote the corresponding test statistic from the data samples. Then, check if the test statistic $W_n$ is in the rejection region $(u,\infty)$ by computing $I_m = \mathbb{I}(W_n > u)$.

The rejection rates for each test statistic are then approximated by $M^{-1}\sum_{m=1}^M I_m$.

Figure 3.17 shows the empirical rejection rates of bootstrap testing when enforcing the null (red) or not (blue) with different sample sizes $n$ and different degrees $p$ of the alternative. The tuning parameters $h_n$ and $g_n$ are selected by our rule of thumb. Meanwhile, Figure 3.18 exhibits the empirical rejection rates of both methods when $g_n$ is selected by our rule of thumb while $h_n$ varies over $\{1,\ldots,15\}$. Here, due to similarity, we only report the results when the sample size is $n=50$, the latent variable $\xi$ follows the standard normal distribution, and the test statistics is of maximum-type.

### 3.11.3 Sampling distributions of non-negligible bias terms

In this Section, we provide extended simulation results as shown in Figure 3.2 of the main paper. We find distributions of non-negligible bias terms in Lemma 34 and Section 3.10.6.3, which are re-defined as follows:

$$B_{\dim,n} \equiv B_{\dim,n}(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle (I-\Pi_{h_n})\Gamma_{g_n}^{-1}U_n, X_0\rangle, \tag{3.95}$$

$$B_{\mathrm{cons},n} \equiv B_{\mathrm{cons},n}(X_0) = \sqrt{\frac{n}{s_{h_n}(X_0)}}\langle \Gamma_{h_n}^{-1}(I-\Pi_{g_n})\hat{\Delta}_n, X_0\rangle. \tag{3.96}$$

Recall from Lemma 34 and Proposition 31 that the non-negligible bias terms $B_{\dim,n}$ and $B_{\mathrm{cons},n}$ weakly converges to normal distributions with asymptotic variances $\sigma^2_{\dim}(\tau) \equiv \tau^{-1}-1$ and $\sigma^2_{\mathrm{cons}}(\tau) \equiv (1-\tau^{-1})\left(\|\Gamma^{1/2}\beta\|^2 \Big/ \sum_{j=1}^\infty \gamma_j\rho_j^2 + 1\right)$, respectively. These results can be also demonstrated numerically in Figures 3.19-3.20. As the forms of limiting variances $\sigma^2_{\dim}(\tau)$ and $\sigma^2_{\mathrm{cons}}(\tau)$ indicate, the distributions of each bias term are more influential and spread more widely, when the ratio $h_n/g_n$ gets far from 1. Due to similarity, we provide only the results when $(n,g_n)=(200,4)$, $(n,g_n)=(1000,6)$, and the error is dependent on the regressor, under the same scenario for the results in Figure 3.3 of the main paper.

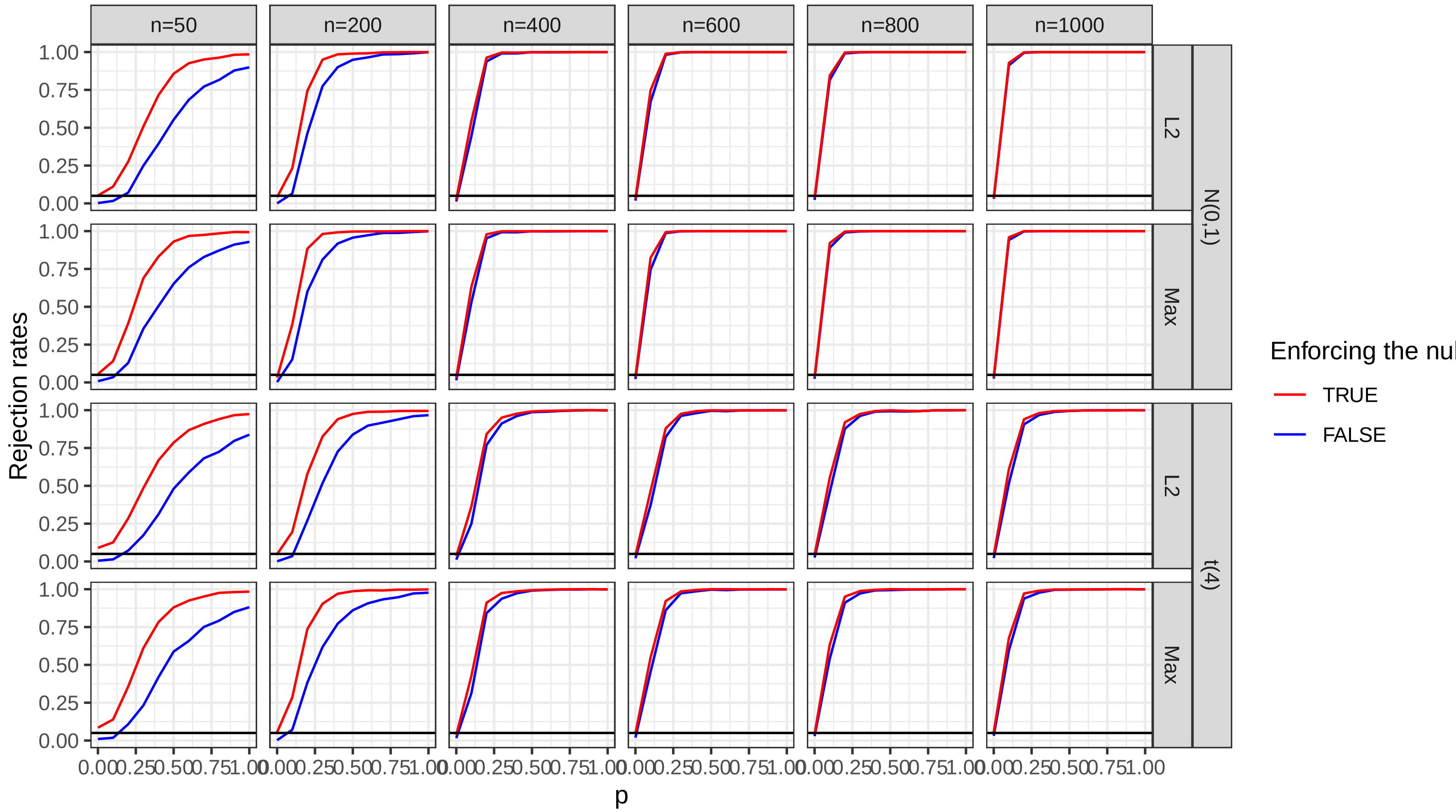


Figure 3.17: Empirical rejection rates of the bootstrap testing procedure for the null hypothesis $H_0 : \Pi_{\mathcal{X}_0^p} \beta = 0$ when enforcing the null (red) or not (blue) with different sample sizes $n$ and different degrees $p$ of the alternative. Here, the latent variable $\xi$ follows either $\mathsf{N}(0, 1)$ or $t(4)$ and the test statistics is either of $L^2$- or maximum-type. The tuning parameters $h_n$ and $g_n$ are selected by our rule of thumb.

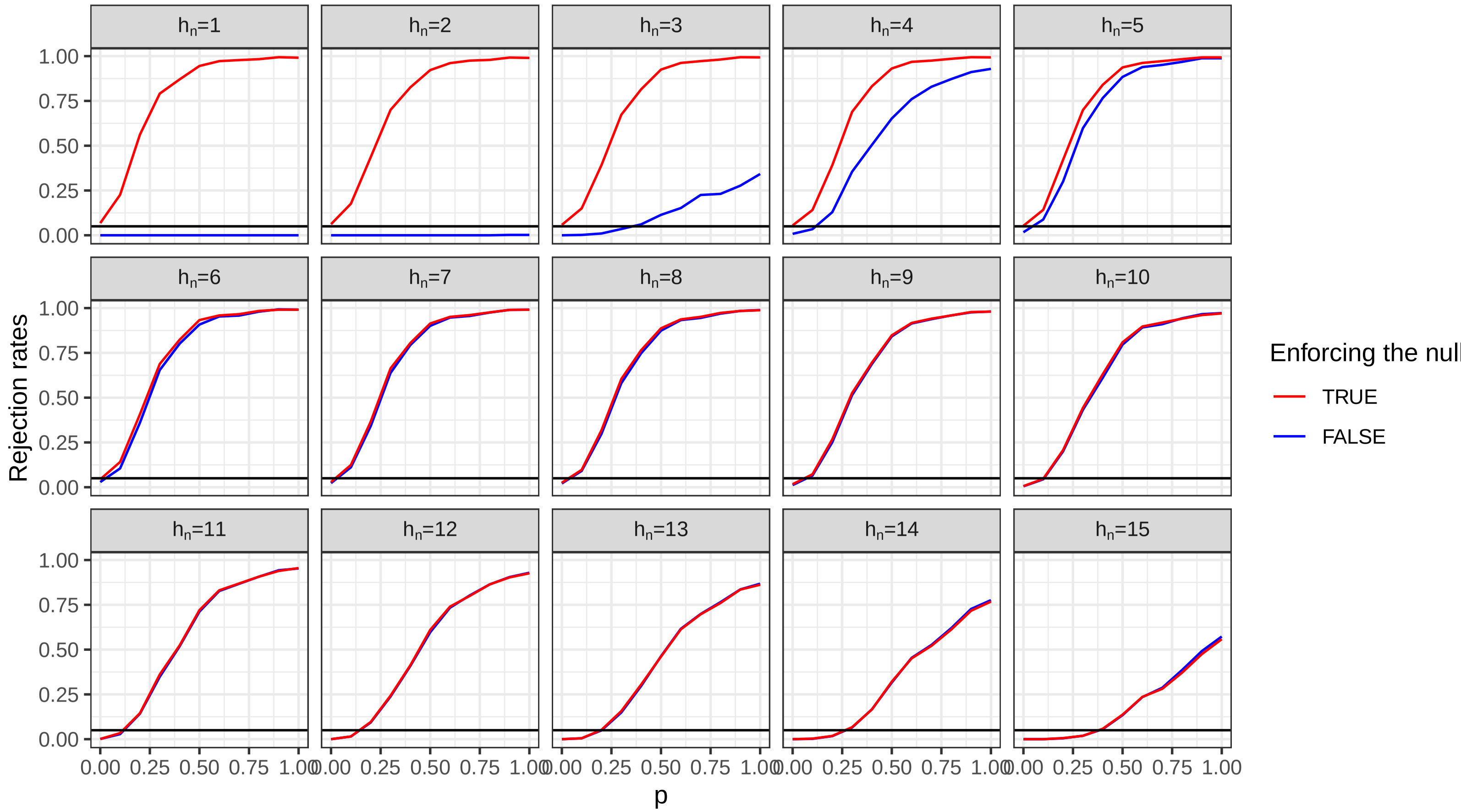


Figure 3.18: Empirical rejection rates of the bootstrap testing procedure for the null hypothesis $H_0 : \Pi_{\mathcal{X}_0^p}\beta = 0$ when enforcing the null (red) or not (blue). Here, the sample size is $n = 50$, the latent variable $\xi$ follows the standard normal distribution, and the test statistics is of maximum-type. The tuning parameter $g_n$ is selected by our rule of thumb while the other tuning parameter $h_n$ varies over $\{1, \ldots, 15\}$.

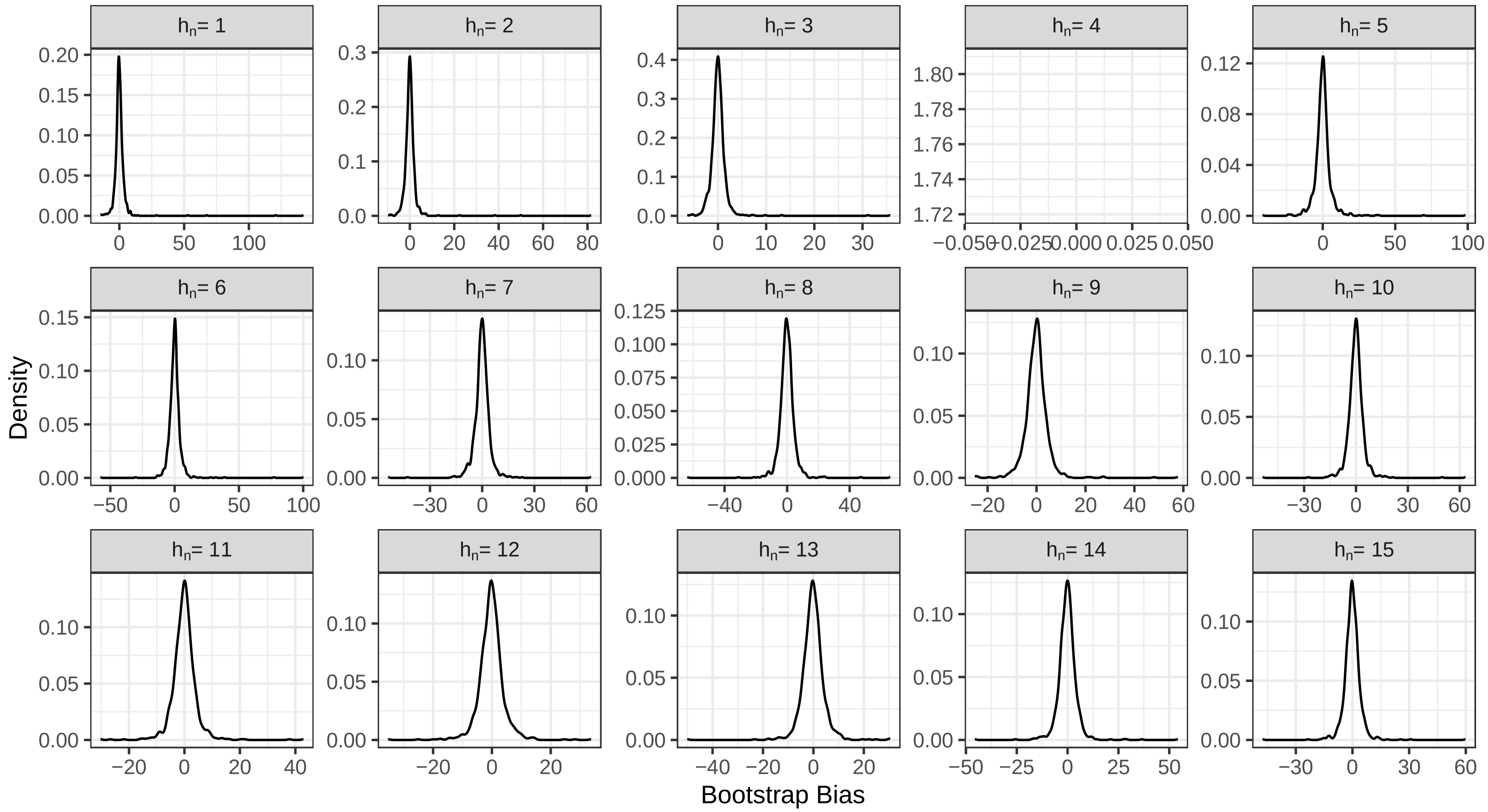


Figure 3.19: Sampling distributions of $B_{\text{dim},n}$ from (3.95) ($h_n < g_n$) and $B_{\text{cons},n}$ from (3.96) ($h_n > g_n$) when $n = 200$ and $g_n = 4$.

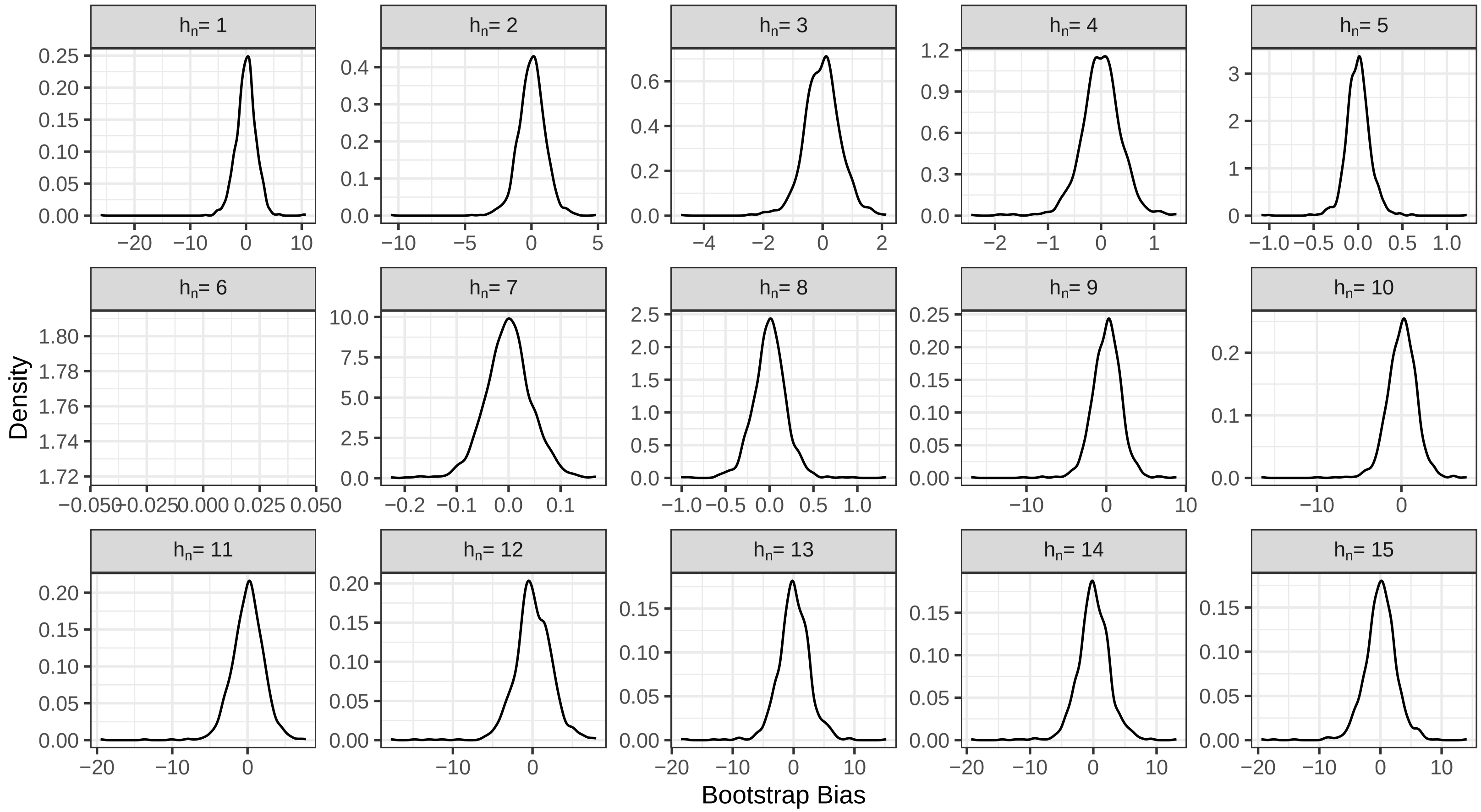


Figure 3.20: Sampling distributions of $B_{\text{dim},n}$ from (3.95) ($h_n < g_n$) and $B_{\text{cons},n}$ from (3.96) ($h_n > g_n$) when $n = 1000$ and $g_n = 6$.

## 3.12 Additional details regarding real data applications

We provide further results from data analysis on Canadian weather dataset from Section 3.6 of the main paper another data application for medfly dataset. For both datasets, we provide the following over different truncation parameters: residual plots, estimated prediction errors from cross-validation, confidence intervals for projections from each bootstrap method, ratios of widths of residual bootstrap intervals to widths of paired bootstrap intervals either with or without studentization, and p-values from the bootstrap testing procedures. The cross-validation method used here is the same as the one described in Section S3.1 of the supplement of [53].

### 3.12.1 Canadian weather dataset

Figure 3.21 shows plots of squared residuals $\hat{\varepsilon}^2_{i,k_n} \equiv n^{-1}\sum_{i=1}^{n}(Y_i^c - \langle\hat{\beta}_{k_n}, X_i^c\rangle)^2$ versus predicted values $\hat{Y}^c_{i,k_n} \equiv \langle\hat{\beta}_{k_n}, X_i^c\rangle$ over different truncation parameters $k_n \in \{1, \ldots, 20\}$.

Figure 3.22 shows estimated prediction errors $\widehat{PE}(k_n)$ (cf. Section S3.1 of the supplement of [53]) of FPCR estimator $\hat{\beta}_{k_n}$ by cross-validation over different truncation parameters $k_n \in \{1, \ldots, 20\}$.

In Figure 3.23, we provide (symmetrized) confidence intervals for (centered) projections $\{\langle\beta, X^c_{0,l}\rangle\}_{l=1}^4$ from each bootstrap method for different truncation parameters $h_n \in \{1, \ldots, 18\}$. The results when $h_n \in \{19, 20\}$ are omitted due to the relatively wide widths of the corresponding confidence intervals. Here, we set $k_n = 2 = g_n$ as described in **??** of the main paper. The plots of width ratios over different truncation parameters $h_n \in \{1, \ldots, 20\}$ are also given in Figure 3.24.

Figure 3.25 displays further p-values from bootstrap testing procedures are provided when truncation parameter $h_n$ varies over $\{1, \ldots, 20\}$.

### 3.12.2 Medfly dataset

#### 3.12.2.1 Main results

We next examine a medfly dataset, as considered in some previous contexts [17, 44, 52, 47]. We adopt a version of the dataset made available in the R package `fdapace`; the full dataset with

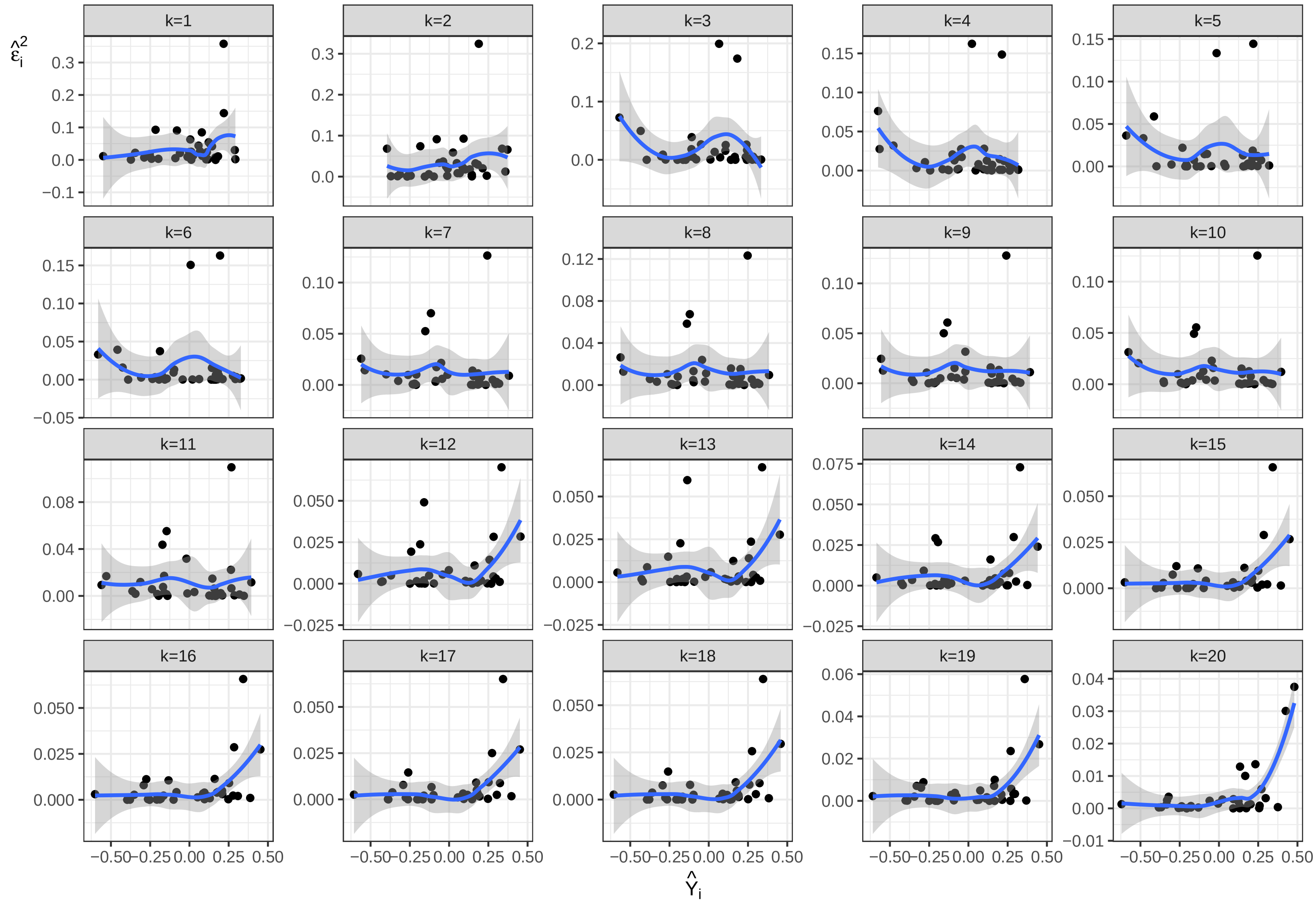


Figure 3.21: Scatterplots of squared residuals $\hat{\varepsilon}^2_{i,k_n}$ versus predicted values $\hat{Y}^c_{i,k_n} \equiv \langle \hat{\beta}_{k_n}, X^c_i \rangle$ over different truncation parameters $k_n \in \{1, \ldots, 20\}$ for Canadian weather dataset.

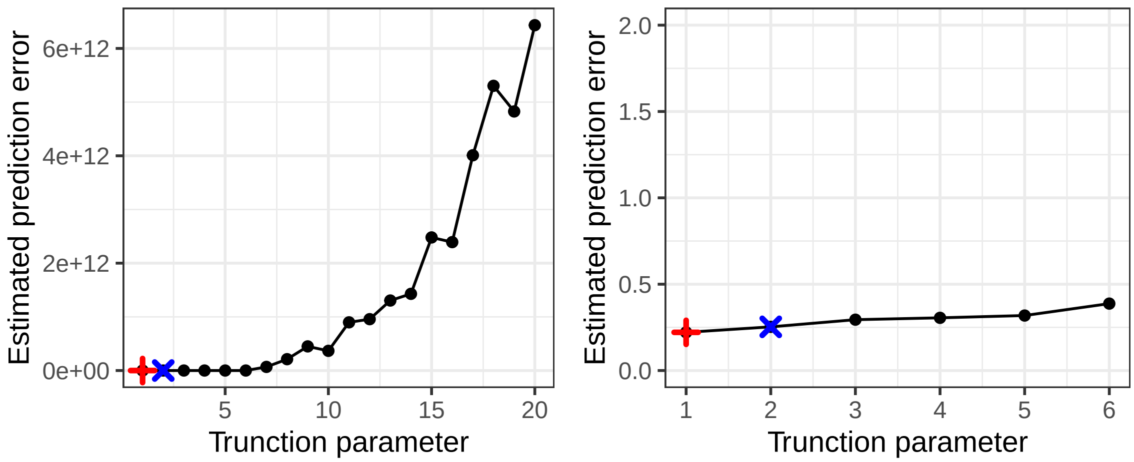


Figure 3.22: Estimated prediction errors $\widehat{PE}(k_n)$ of FPCR estimator $\hat{\beta}_{k_n}$ by cross-validation over different truncation parameters $k_n \in \{1, \ldots, 20\}$ for Canadian weather dataset.

its experimental background is described in [13]. For 789 female Mediterranean fruit flies (medflies, *Ceratitis capitata*), each regressor curve $X_i$ represents daily measurements of the number of eggs laid by a medfly in the first 25 days of life while the response $Y_i$ is the total remaining number of eggs laid during a lifetime. Randomly selected subsamples with different sample sizes $n \in \{50, 150, \ldots, 650, 789\}$ are also chosen to study the effect of the sample size on the bootstrap intervals.

We investigate whether the variability in the responses might depend on the type of start to a medfly's egg-laying, as potentially suggesting heteroscedastic errors in a FLRM (3.2). To do this, we classify a medfly as a slow starter if it lays a first egg after ten days; otherwise, the medfly is called an early starter. Ten representative curves for both early and slow starters are provided in Figure 3.26 along with average curves from the whole data. To check for heteroscedasticity, we examine the estimated standard deviation for either early or slow starters, which is provided in Figure 3.27 for various subsampled datasets. This shows that the homoscedastic error assumption appears inappropriate in general, in which case only the PB method would be applicable.

Similar to previous analyses of these data, we use average curves for either early and slow starters as new predictors. We add also two additional predictor curves based on [44], who

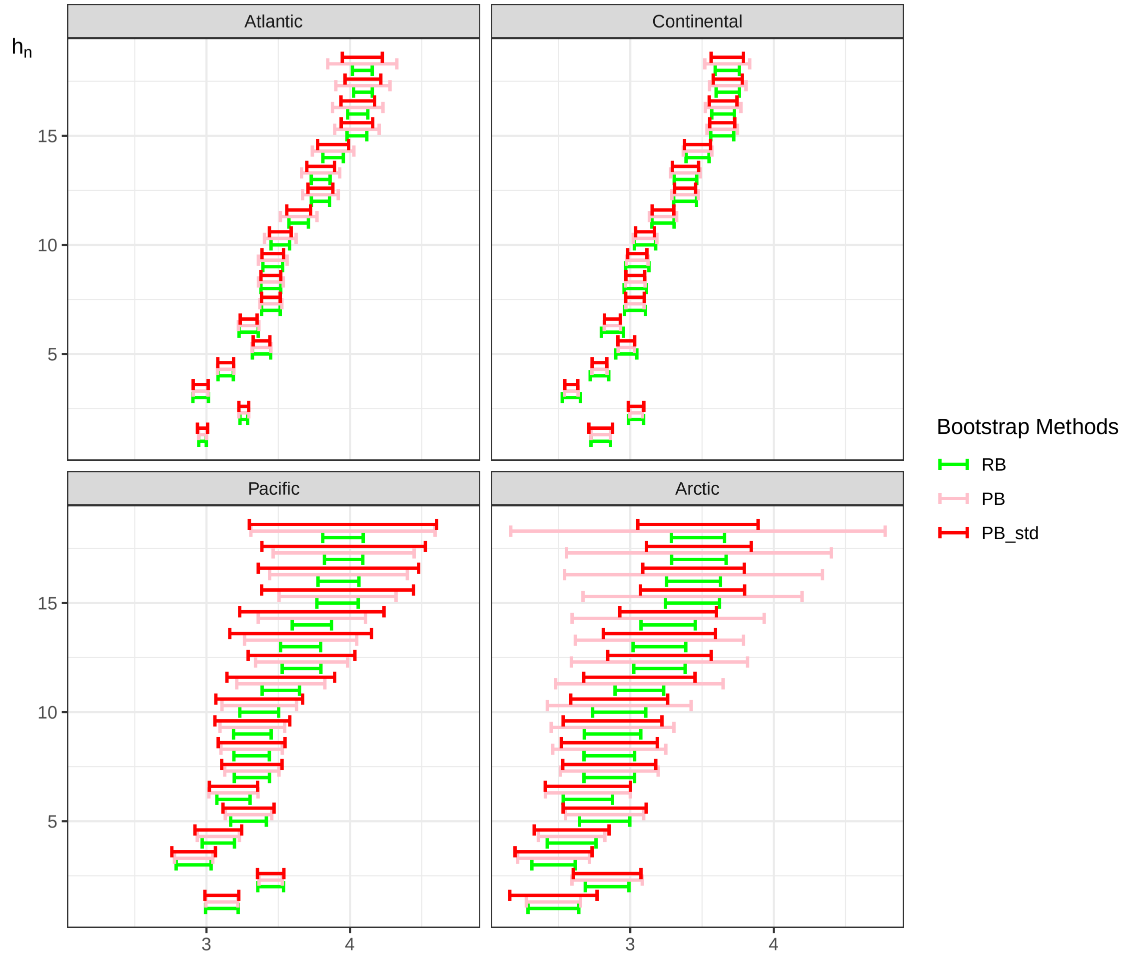


Figure 3.23: Symmetrized confidence intervals from each bootstrap method over different truncation parameters $h_n \in \{1, \ldots, 18\}$ with $k_n = 2 = g_n$ for Canadian weather dataset.

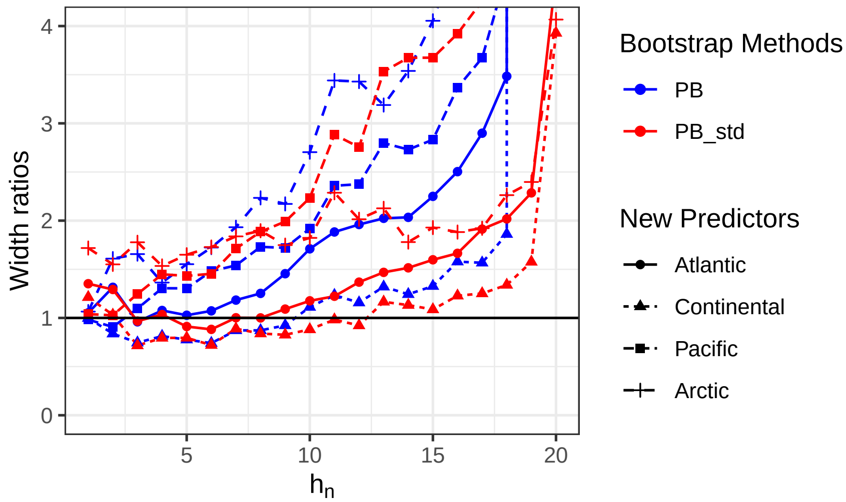


Figure 3.24: Ratios of widths of residual bootstrap intervals to widths of paired bootstrap intervals either with or without studentization over different truncation parameters $h_n \in \{1, \ldots, 20\}$ with $k_n = 2 = g_n$ for Canadian weather dataset.

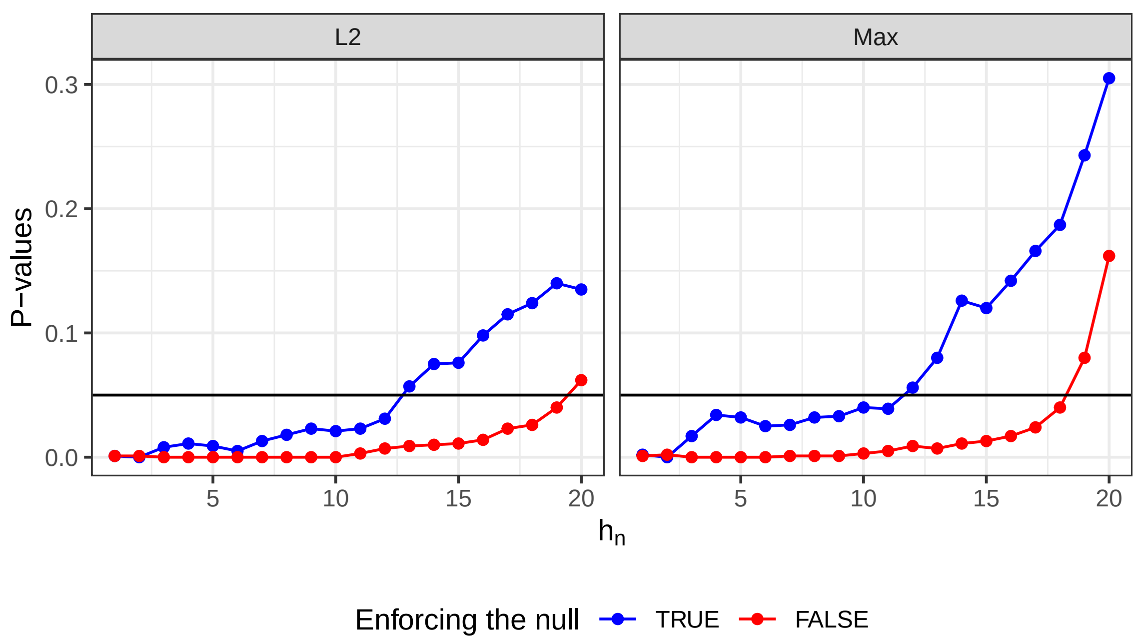


Figure 3.25: P-values from bootstrap testing procedures over different truncation parameters $h_n \in \{1, \ldots, 20\}$ with $k_n = 2 = g_n$ for Canadian weather dataset.

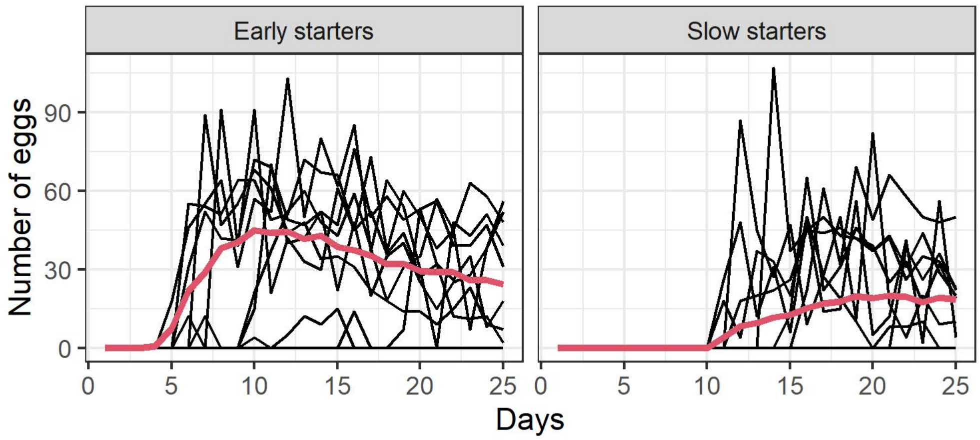


Figure 3.26: Randomly selected ten observed curves of medfly dataset in each group of either early or slow starters with their average curves denoted in thicker pink lines.

classify the medfly curves into two groups: long- and short-lived medflies. This gives four regressor curves, given in Figure 3.28, determined by the groupwise averages. At these predictors, (symmetrized) confidence intervals for mean responses are provided in Figure 3.29 for different subsample sizes and bootstrap methods; as in Section 3.6, the data observations and new predictors are centered by the averages of each subsample before analysis and we consider RB as well as two PB versions (PB, PB_std). Interval widths depend on the predictor curve (medfly group), though these decrease as the sample size increases. Slow starters have intervals with the widest ranges, including the highest and lowest values of total number of eggs laid, while intervals for early starters intervals are located around small egg totals, relative to the overall average. The intervals for long- and short-lived group predictors possess relatively narrow widths and center on closer to zero.

Figure 3.30 shows the results from bootstrap tests of whether the projections of the slope function onto the space spanned by four (centered) predictors $\mathcal{X}_0^c \equiv \{X_{0,l}^c\}_{l=1}^4$ may be zero. The maximum type test statistic from (3.20) is reported, as the other test form is similar. Bootstrap tests conclusively reject the null hypothesis, which strongly supports that the slope function is not orthogonal to the spanned by the predictors under consideration.

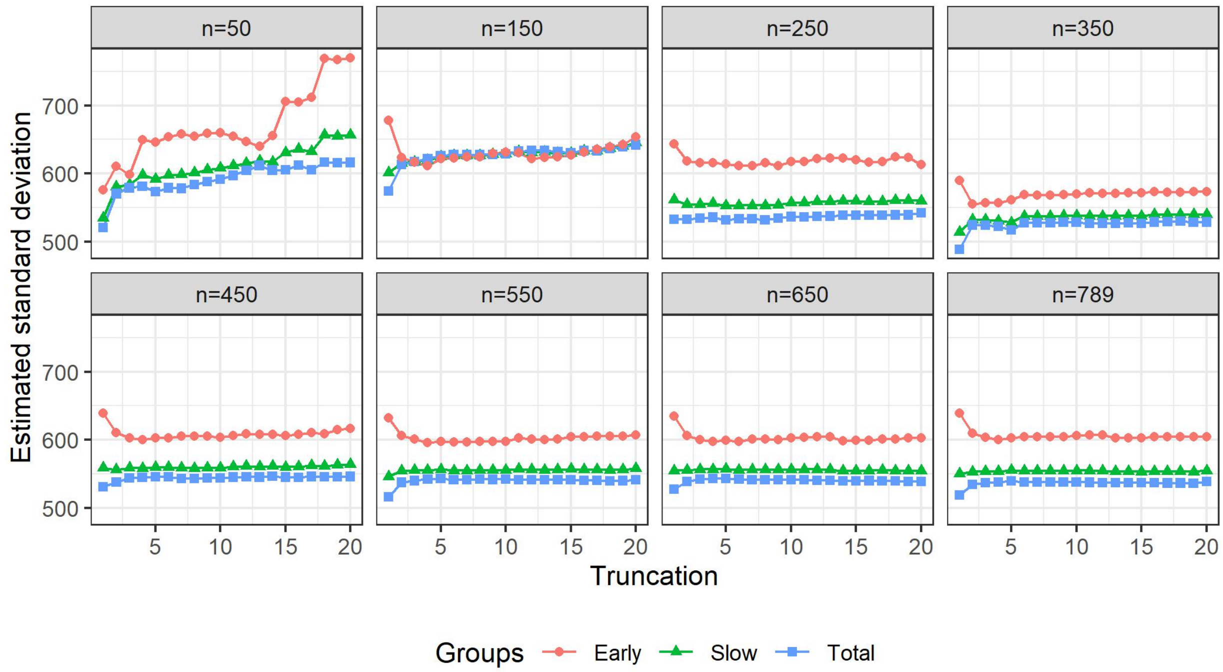


Figure 3.27: Estimated standard deviations of either early or slow starter groups over different truncation levels for medfly dataset.

#### 3.12.2.2 Extra results

We provide the same plots for medfly dataset as shown in Section 3.12.1.

Figure 3.31 shows plots of squared residuals $\hat{\varepsilon}^2_{i,k_n}$ versus predicted values $\hat{Y}_{i,k_n} \equiv \langle \hat{\beta}_{k_n}, X_i \rangle$ over different truncation parameters $k_n \in \{1, \ldots, 5\}$.

Figure 3.32 shows estimated prediction errors $\widehat{PE}(k_n)$ (cf. Section S3.1 of the supplement of [53]) of FPCR estimator $\hat{\beta}_{k_n}$ by cross-validation over different truncation parameters $k_n \in \{1, \ldots, 20\}$.

In Figure 3.33, we provide confidence intervals for projections $\{\langle \beta, X^c_{0,l} \rangle\}_{l=1}^4$ from each bootstrap method for different truncation parameters $h_n \in \{1, \ldots, 20\}$. Here, $k_n = g_n$ are chosen as the truncation level that gives the minimum prediction error (cf. Figure 3.32) for each subsample size. The plots of width ratios over different truncation parameters $h_n \in \{1, \ldots, 20\}$ are also given in Figure 3.34.

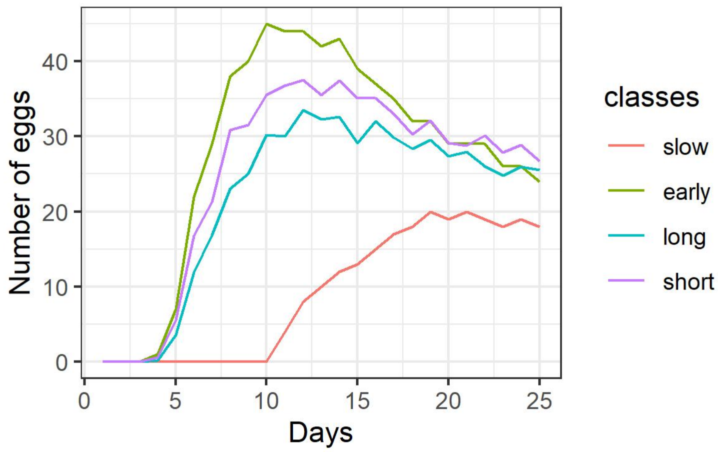


Figure 3.28: New regressor curves under considerations for medfly dataset.

Figure 3.35 displays further p-values from max-type bootstrap testing procedures are provided when truncation parameter $h_n$ varies over $\{1, \ldots, 20\}$.

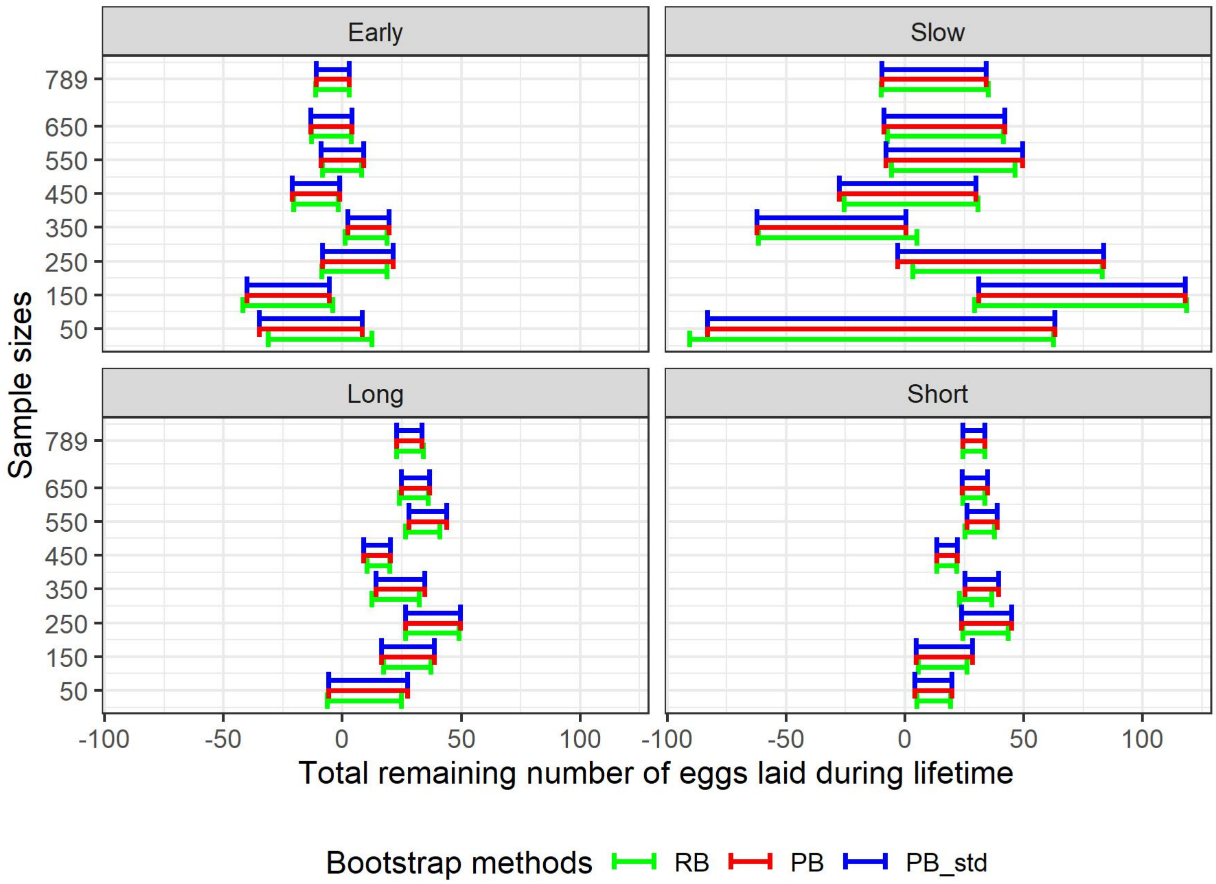


Figure 3.29: Bootstrap confidence intervals for projections $\{\langle \beta, X_{0,l}^{c} \rangle\}_{l=1}^{4}$ from each subsample in medfly dataset.

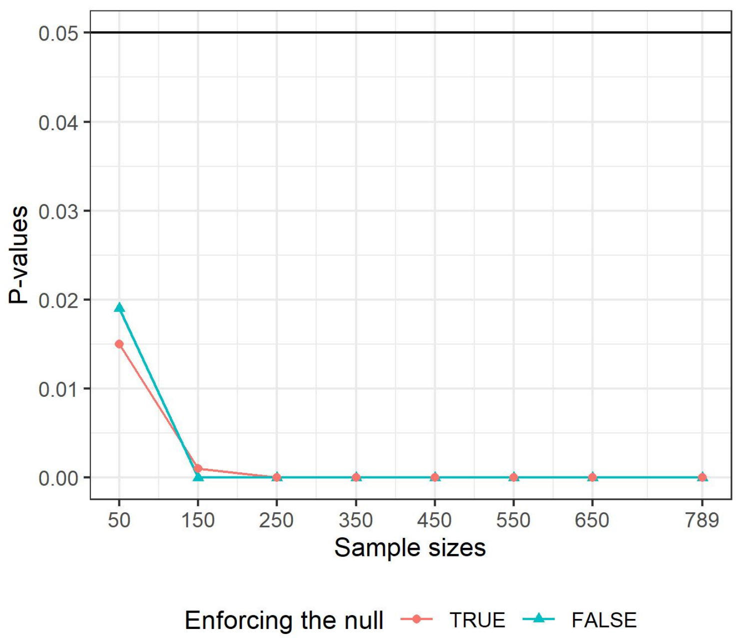


Figure 3.30: P-values from bootstrap testing of the null hypothesis $H_0 : \Pi_{\mathcal{X}_0^c}\beta = 0$ with different sample size in medfly dataset, where the maximum type statistics are used.

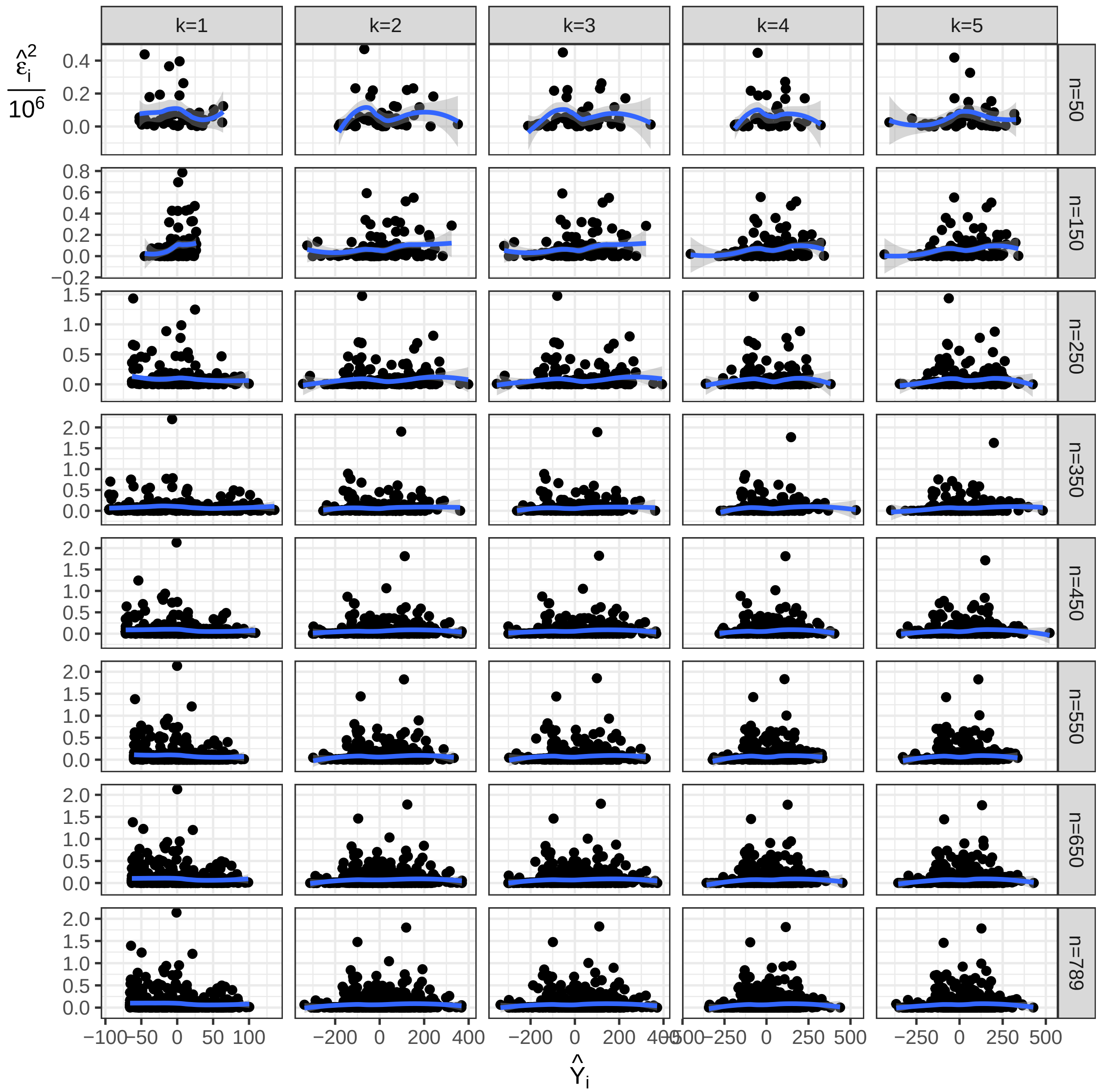


Figure 3.31: Scatterplots of squared residuals $\hat{\varepsilon}^2_{i,k_n}$ versus predicted values $\hat{Y}^c_{i,k_n} \equiv \langle \hat{\beta}_{k_n}, X^c_i \rangle$ over different truncation parameters $k_n \in \{1, \ldots, 20\}$ for subsamples of medfly dataset.

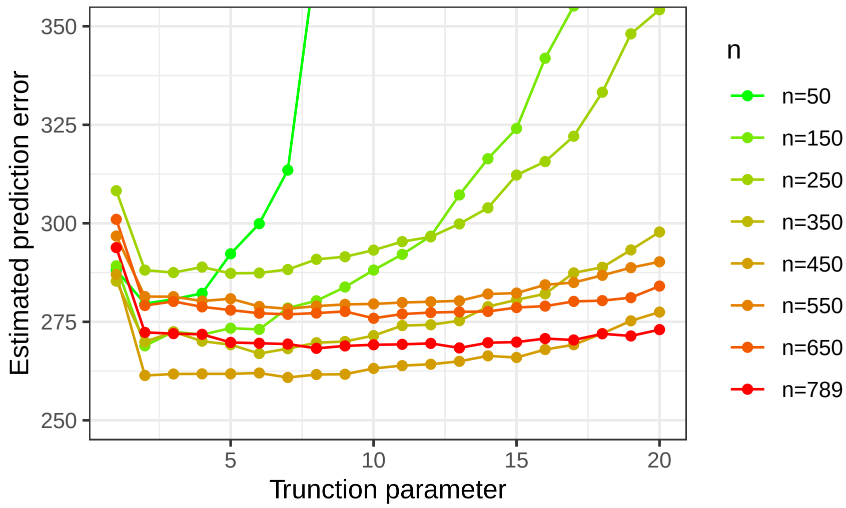


Figure 3.32: Estimated prediction errors $\widehat{PE}(k_n)$ of FPCR estimator $\hat{\beta}_{k_n}$ by cross-validation over different truncation parameters $k_n \in \{1, \ldots, 20\}$ for subsamples of medfly dataset.

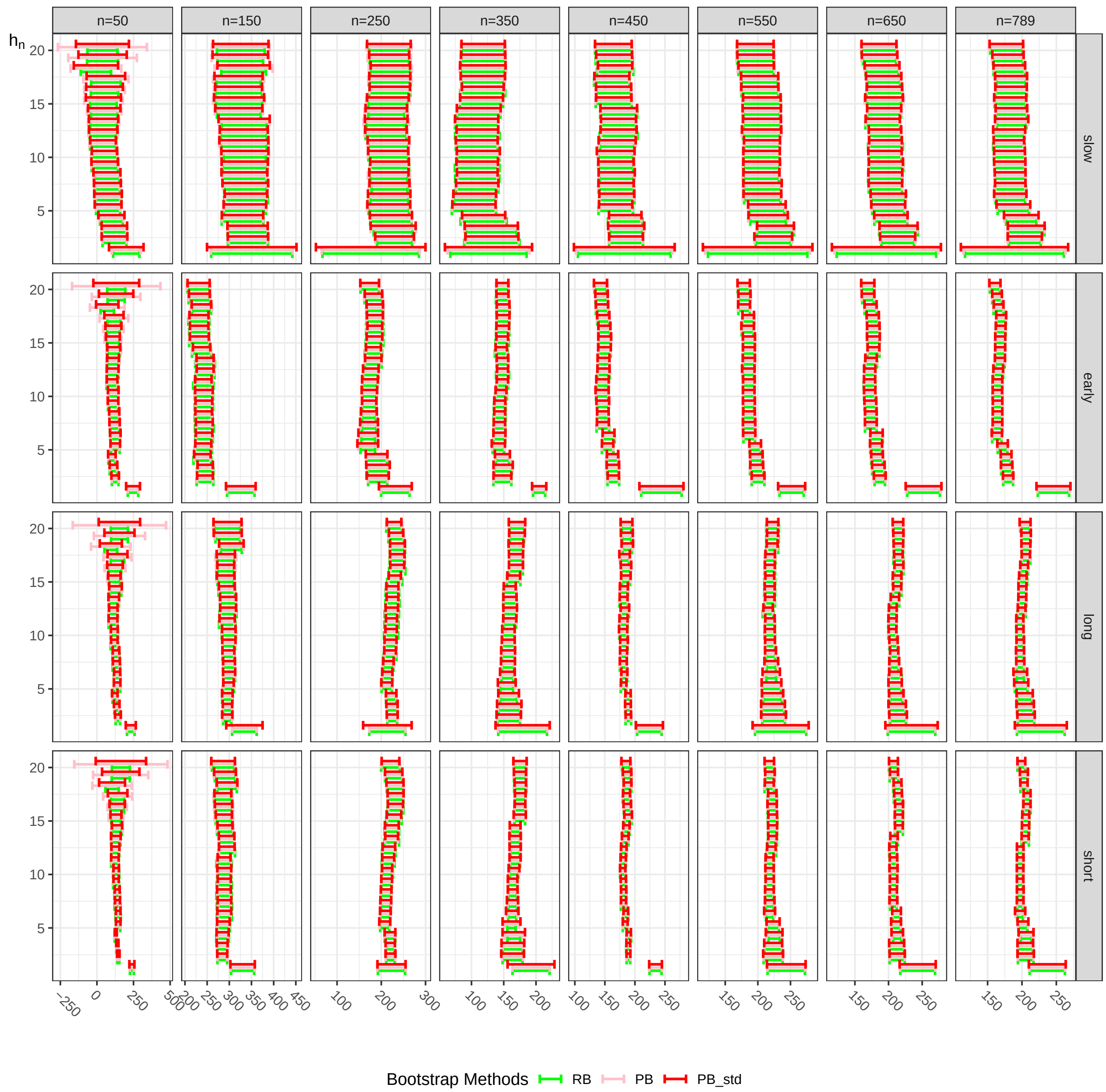


Figure 3.33: Confidence intervals from each bootstrap method over different truncation parameters $h_n \in \{1, \ldots, 20\}$ with $g_n = k_n$ for subsamples of medfly dataset.

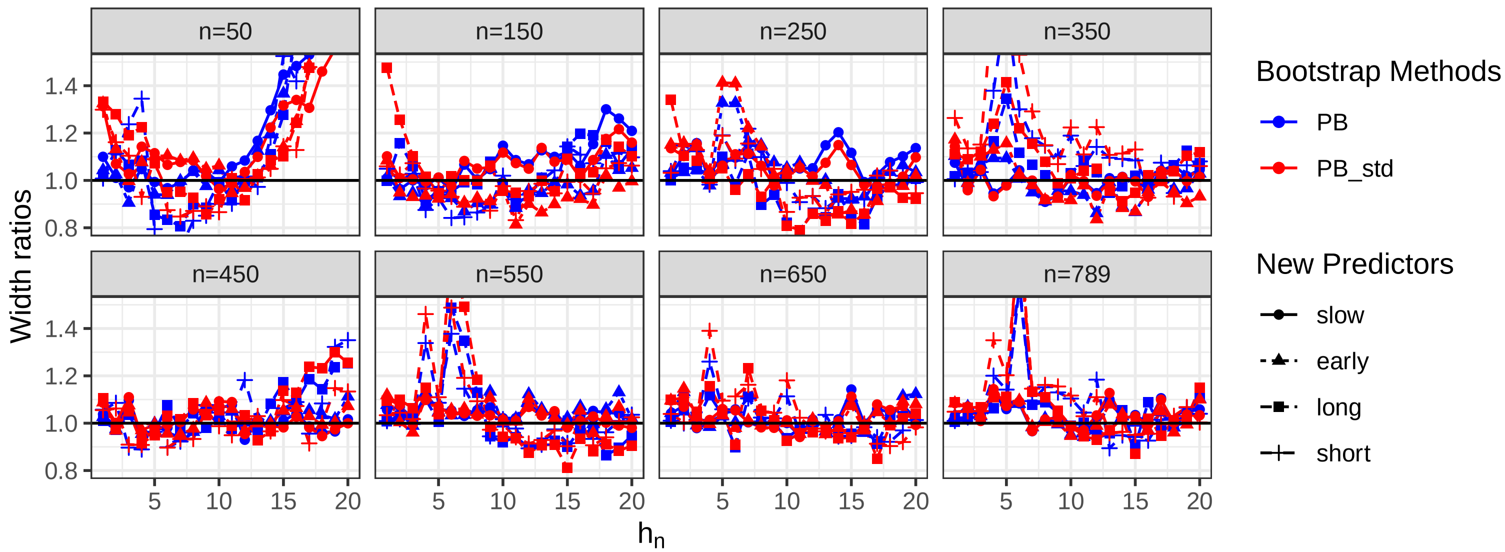


Figure 3.34: Ratios of widths of residual bootstrap intervals to widths of paired bootstrap intervals either with or without studentization over different truncation parameters $h_n \in \{1, \ldots, 20\}$ with $g_n = k_n$ for subsamples of medfly dataset.

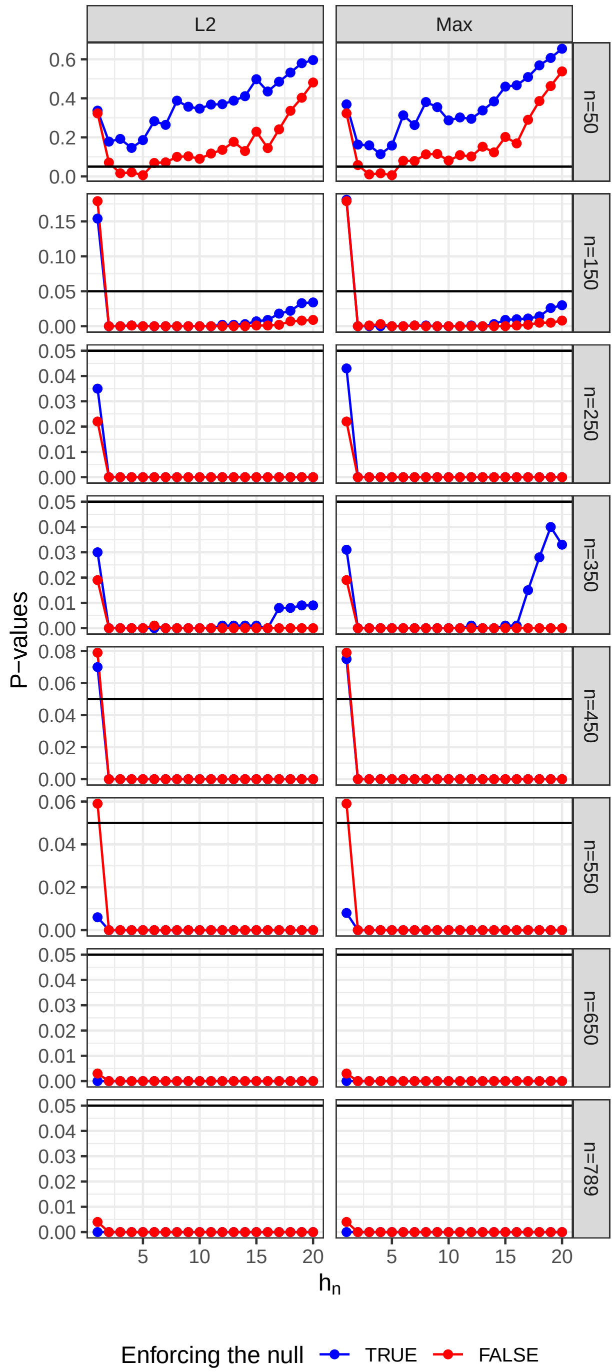


Figure 3.35: P-values from max-type bootstrap testing procedures over different truncation parameters $h_n \in \{1, \dots, 20\}$ with $g_n = k_n$ for subsamples of medfly dataset.

### 3.13 References

[1] Ahlfors, L. *Complex Analysis: An Introduction to The Theory of Analytic Functions of One Complex Variable*. International series in pure and applied mathematics. McGraw-Hill Education, 1979.

[2] Athreya, K. B., and Lahiri, S. N. *Measure Theory and Probability Theory*. Springer Science & Business Media, 2006.

[3] Bickel, P. J., and Freedman, D. A. Some asymptotic theory for the bootstrap. *The annals of statistics* (1981), 1196–1217.

[4] Billingsley, P. *Probability and Measure*. John Wiley & Sons, 1999.

[5] Cai, T. T., and Hall, P. Prediction in functional linear regression. *The Annals of Statistics 34*, 5 (2006), 2159–2179.

[6] Cai, T. T., and Yuan, M. Minimax and Adaptive Prediction for Functional Linear Regression. *Journal of the American Statistical Association 107*, 499 (Sept. 2012), 1201–1216.

[7] Camponovo, L. On the validity of the pairs bootstrap for lasso estimators. *Biometrika 102*, 4 (2015), 981–987.

[8] Cardot, H., Ferraty, F., Mas, A., and Sarda, P. Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics 30*, 1 (2003), 241–255.

[9] Cardot, H., Ferraty, F., and Sarda, P. Functional linear model. *Statistics & Probability Letters 45*, 1 (1999), 11–22.

[10] Cardot, H., Ferraty, F., and Sarda, P. Spline estimators for the functional linear model. *Statistica Sinica* (2003), 571–591.

[11] Cardot, H., Goia, A., and Sarda, P. Testing for no effect in functional linear regression models, some computational approaches. *Communications in Statistics-Simulation and Computation 33*, 1 (2004), 179–199.

[12] Cardot, H., Mas, A., and Sarda, P. Clt in functional linear regression models. *Probability Theory and Related Fields 138*, 3-4 (2007), 325–361.

[13] Carey, J. R., Liedo, P., Müller, H.-G., Wang, J.-L., and Chiou, J.-M. Relationship of age patterns of fecundity to mortality, longevity, and lifetime reproduction in a large cohort of mediterranean fruit fly females. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 53*, 4 (1998), B245–B251.

[14] Chatterjee, A., and Lahiri, S. N. Asymptotic properties of the residual bootstrap for lasso estimators. *Proceedings of the American Mathematical Society 138*, 12 (2010), 4497–4509.

[15] Chatterjee, A., and Lahiri, S. N. Bootstrapping lasso estimators. *Journal of the American Statistical Association 106*, 494 (2011), 608–625.

[16] Cheng, G., and Chen, Y.-C. Nonparametric inference via bootstrapping the debiased estimator. *Electronic Journal of Statistics 13*, 1 (2019), 2194–2256.

[17] Chiou, J.-M., Müller, H.-G., and Wang, J.-L. Functional quasi-likelihood regression models with smooth random effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*, 2 (2003), 405–423.

[18] Crambes, C., and Mas, A. Asymptotics of prediction in functional linear regression with functional outputs. *Bernoulli 19*, 5B (2013), 2627–2651.

[19] Delaigle, A., Hall, P., and Apanasovich, T. V. Weighted least squares methods for prediction in the functional data linear model. *Electronic Journal of Statistics 3* (2009), 865–885.

[20] Dezeure, R., Bühlmann, P., and Zhang, C.-H. High-dimensional simultaneous inference with the bootstrap. *Test 26*, 4 (2017), 685–719.

[21] Dunford, N., and Schwartz, J. T. *Linear Operators, Part 1: General Theory*, vol. 10. John Wiley & Sons, 1988.

[22] Freedman, D. A. Bootstrapping regression models. *The Annals of Statistics 9*, 6 (1981), 1218–1228.

[23] Gohberg, I., Goldberg, S., and Kaashoek, M. *Classes of Linear Operators Vol. I*, vol. 49. Birkhäuser, 1990.

[24] González-Manteiga, W., González-Rodríguez, G., Martínez-Calvo, A., and García-Portugués, E. Bootstrap independence test for functional linear models. *arXiv preprint arXiv:1210.1072* (2012).

[25] González-Manteiga, W., and Martínez-Calvo, A. Bootstrap in functional linear regression. *Journal of Statistical Planning and Inference 141*, 1 (2011), 453–461.

[26] Hall, P. Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *The Annals of Statistics* (1992a), 675–694.

[27] Hall, P. On bootstrap confidence intervals in nonparametric regression. *The Annals of Statistics* (1992b), 695–711.

[28] Hall, P., Horowitz, J. L., et al. Methodology and convergence rates for functional linear regression. *Annals of Statistics 35*, 1 (2007), 70–91.

[29] Härdle, W., and Mammen, E. Bootstrap methods in nonparametric regression. In *Nonparametric Functional Estimation and Related Topics*. Springer, 1991, pp. 111–123.

[30] Hardle, W., and Mammen, E. Comparing nonparametric versus parametric regression fits. *The Annals of Statistics 21*, 4 (1993), 1926–1947.

[31] Hilgert, N., Mas, A., and Verzelen, N. Minimax adaptive tests for the functional linear model. *The Annals of Statistics 41*, 2 (2013), 838–869.

[32] Horváth, L., and Kokoszka, P. *Inference for Functional Data with Applications*, vol. 200. Springer Science & Business Media, 2012.

[33] Hsing, T., and Eubank, R. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, vol. 997. John Wiley & Sons, 2015.

[34] Khademnoe, O., and Hosseini-Nasab, S. M. E. On asymptotic distribution of prediction in functional linear regression. *Statistics 50*, 5 (2016a), 974–990.

[35] Khademnoe, O., and Hosseini-Nasab, S. M. E. On properties of percentile bootstrap confidence intervals for prediction in functional linear regression. *Journal of Statistical Planning and Inference 170* (2016b), 129–143.

[36] Kong, D., Staicu, A.-M., and Maity, A. Classical testing in functional linear models. *Journal of nonparametric statistics 28*, 4 (2016), 813–838.

[37] Koul, H. L., and Lahiri, S. N. On bootstrapping m-estimated residual processes in multiple linear-regression models. *Journal of Multivariate Analysis 49*, 2 (1994), 255–265.

[38] Lahiri, S. *Resampling Methods for Dependent Data*. Springer Science & Business Media, 2003.

[39] Lahiri, S. N. Bootstrapping $m$-estimators of a multiple linear regression parameter. *The Annals of Statistics 20*, 3 (1992), 1548–1570.

[40] Lei, J. Adaptive global testing for functional linear models. *Journal of the American Statistical Association 109*, 506 (2014), 624–634.

[41] Lin, Y., and Lin, Z. A unified approach to hypothesis testing for functional linear models. *arXiv preprint arXiv:2109.02309* (2021).

[42] Mammen, E. Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics 21*, 1 (1993), 255–285.

[43] Müller, H.-G., Chiou, J.-M., and Leng, X. Inferring gene expression dynamics via functional regression analysis. *BMC bioinformatics 9*, 1 (2008), 1–20.

[44] Müller, H.-G., and Stadtmüller, U. Generalized functional linear models. *The Annals of Statistics 33*, 2 (2005), 774–805.

[45] Ramsay, J., and Silverman, B. *Functional Data Analysis*. Springer. Springer, 2005.

[46] Reiss, P. T., and Ogden, R. T. Functional generalized linear models with images as predictors. *Biometrics 66*, 1 (2010), 61–69.

[47] Sguera, C., and López-Pintado, S. A notion of depth for sparse functional data. *Test 30*, 3 (2021), 630–649.

[48] Shorack, G. R. Bootstrapping robust regression. *Communications in statistics-theory and methods 11*, 9 (1982), 961–972.

[49] Su, Y.-R., Di, C.-Z., and Hsu, L. Hypothesis testing in functional linear models. *Biometrics 73*, 2 (2017), 551–561.

[50] Wang, H. J., McKeague, I. W., and Qian, M. Testing for marginal linear effects in quantile regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*, 2 (2018), 433–452.

[51] Wang, X., Zhu, H., and Initiative, A. D. N. Generalized scalar-on-image regression models via total variation. *Journal of the American Statistical Association 112*, 519 (2017), 1156–1168.

[52] Wu, Y., Fan, J., and Müller, H.-G. Varying-coefficient functional linear regression. *Bernoulli 16*, 3 (2010), 730–758.

[53] Yeon, H., Dai, X., and Nordman, D. J. Bootstrap inference in functional linear regression models with scalar response. *accepted by Bernoulli Journal* (2023).

[54] Yeon, H., Dai, X., and Nordman, D. J. Supplement to "bootstrap inference in functional linear regression models with scalar response under heteroscedasticity".

[55] Ziegler, K. On bootstrapping the mode in the nonparametric regression model with random design. *Metrika 53*, 2 (2001), 141–170.

# CHAPTER 4. INITIAL THEORETICAL WORK ON WILD BOOTSTRAP FOR FUNCTIONAL LINEAR REGRESSION

Modified from a working manuscript to be submitted to Statistical Science

Hyemin Yeon[1], Xiongtao Dai[2], and Daniel John Nordman[1]

[1]Department of Statistics, Iowa State University

[2]Division of Biostatistics, University of California, Berkeley

**Abstract**

This paper is an initial theoretical work on wild bootstrap for functional linear regression. In functional linear regression model, the inference under heteroscedastic error assumptions have not received much attention. We propose a multiplier wlid bootstrap method to approximate sampling distributions, which is expected to be computationally fast. Its theoretical validity is provided under mild assumptions.

## 4.1 Introduction

Researchers have paid attention to the inference of the slope function in functional linear regression models (FLRMs). A common approach is to use an estimator based on functional principal component analysis (FPCA) (cf. Section 2). With such estimators, central limit theorems (CLTs) have been established [6, 18, 19], which are the foundation for asymptotic inference. Based on some CLT, [9, 18, 19] have developed different bootstrap methods in FLRMs.

Despite of increasing attention, inference in heteroscedastic FLRMs has been rarely studied. To the best of our knowledge, only [19] considers inference in FLRMs, where the conditional variances of the errors are heterogeneous, who proposes a paired bootstrap (PB) to incorporate such heteroscedastic errors. However, in large samples, a PB method may become computationally burdensome due to repeated computations of (pseudo-)inverses in estimation. This challenge motivates us to develop alternative resampling method for large data cases.

Wild bootstrap has been suggested as an alternative to paired bootstrap in classical regression problems [17, 14, 10, 15, 8, 9, 7], but not in FLRMs under heteroscedasticity. We propose a wild bootstrap method for the FLRM setting, particularly by using multipliers. In principle, the bootstrap errors in wild bootstrap mimic the true errors by equating their moments and the corresponding powers of residuals (cf. Section 3). The proposed wild bootstrap is theoretical valid to approximate the sampling distribution.

This paper is organized as follows. In Section 4.2, we provide a brief overview of FLRMs and an estimation approach based on FPCA. Section 4.3 then describes the procedure of the proposed wild bootstrap and its theoretical validity. Section 4.4 finally devotes theoretical details.

## 4.2 Functional linear regression models

We consider the following FLRM with scalar response

$$Y = \langle \beta, X \rangle + \varepsilon, \tag{4.1}$$

where $Y$ is a scalar response; $X$ is a functional regressor; and $\beta$ is the slope function. Here, the slope function is assumed to lie in a separable Hilbert space $\mathbb{H}$ with inner product $\langle \cdot, \cdot \rangle$ and $X$ is a random function that take values in $\mathbb{H}$. The error $\varepsilon$ is commonly assumed to have zero mean as $\mathsf{E}[\varepsilon | X] = 0$. Without loss of generality, we assume that $\mathsf{E}[X] = 0$ and $\mathsf{E}[Y] = 0$. Write $\Gamma \equiv \mathsf{E}[X \otimes X]$ and $\Delta \equiv \mathsf{E}[XY]$ for the covariance operator of $X$ and the cross-covariance function of $X$ and $Y$, respectively. Here, $\otimes$ denotes the tensor product between two elements in $\mathbb{H}$, which is defined as $(x \otimes y)(z) = \langle z, x \rangle y$ for $x, y, z \in \mathbb{H}$. The normal equations for FLRM (4.1) is written as

$$\Delta = \Gamma \beta. \tag{4.2}$$

Then, the slope function is given as $\beta = \Gamma^{-1} \Delta$ under the following assumption, which justifies the model identifiability [4, 5, 6]:

(M) $\ker \Gamma = \{0\}$.

For estimation, we suppose $n$ pairs $\{(X_i, Y_i)\}_{i=1}^n$ are randomly generated from the FLRM (4.1), namely,

$$Y_i = \langle \beta, X_i \rangle + \varepsilon_i, \quad i = 1, \dots, n. \tag{4.3}$$

The sample versions of $\Gamma$ and $\Delta$ are defined as $\hat{\Gamma}_n \equiv n^{-1} \sum_{i=1}^n (X_i - \bar{X})^{\otimes 2}$ and $\hat{\Delta}_n \equiv n^{-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$, where $\bar{X} \equiv n^{-1} \sum_{i=1}^n X_i$, $\bar{Y} \equiv n^{-1} \sum_{i=1}^n Y_i$, and $x^{\otimes 2} \equiv x \otimes x$ for $x \in \mathbb{H}$. The sample covariance operator $\hat{\Gamma}_n$ admits spectral decomposition $\hat{\Gamma}_n = \sum_{j=1}^n \hat{\gamma}_j (\hat{\phi}_j \otimes \hat{\phi}_j)$, where $\hat{\gamma}_j \geq 0$ is the $j$-th sample eigenvalue and $\hat{\phi}_j \in \mathbb{H}$ is the corresponding eigenfunction. By regularizing the inversion of $\hat{\Gamma}_n$, a regression estimator $\hat{\beta}_{h_n}$ of $\beta$ is defined as

$$\hat{\beta}_{h_n} \equiv \hat{\Gamma}_{h_n}^{-1} \hat{\Delta}_n, \tag{4.4}$$

where $\hat{\Gamma}_{h_n}^{-1} \equiv \sum_{j=1}^{h_n} \hat{\gamma}_j^{-1} (\hat{\phi}_j \otimes \hat{\phi}_j)$ is a pseudo-inverse of $\hat{\Gamma}_n$. Here, the truncation level $h_n$ represents the number of eigenpairs used in estimation.

Let $X_0$ denote a new regressor under the model, which is independent of $\{(X_i, Y_i)\}_{i=1}^n$. For an observed or a given value of $X_0$, we consider the sampling distribution of the projection statistic

$$T_n(X_0) \equiv \sqrt{\frac{n}{s_{h_n}(X_0)}} [\langle \hat{\beta}_{h_n}, X_0 \rangle - \langle \beta, X_0 \rangle]. \tag{4.5}$$

Here, the scaling factor $s_{h_n}(X_0)$ is defined as

$$s_{h_n}(x) \equiv \langle \Lambda \Gamma_{h_n}^{-1} x, \Gamma_{h_n}^{-1} x \rangle, \quad x \in \mathbb{H},$$

where $\Lambda \equiv \mathsf{E}[(X\varepsilon)^{\otimes 2}]$ is the covariance operator of $X\varepsilon$ and $\Gamma_{h_n}^{-1} \equiv \sum_{j=1}^{h_n} \gamma_j^{-1} (\phi_j \otimes \phi_j)$ denotes a truncated inverse of $\Gamma$ based on the eigenpairs $\{(\gamma_j, \phi_j)\}_{j=1}^\infty$ of $\Gamma$.

### 4.3 Wild bootstrap

To implement the wild bootstrap (WB), we consider another tuning parameter $k_n$ to construct residuals $\hat{\varepsilon}_i \equiv Y_i - \langle \hat{\beta}_{k_n}, X_i \rangle$, $i = 1, \dots, n$, from the estimator $\hat{\beta}_{k_n} \equiv \hat{\Gamma}_{k_n}^{-1} \hat{\Delta}_n$ akin to

(4.4). We then define bootstrap errors $\{\varepsilon_i^*\}_{i=1}^n$ independently drawn from an arbitrary distribution with $\mathsf{E}^*[\varepsilon_i^*] = 0$ and $\mathsf{E}^*[(\varepsilon_i^*)^2] = \hat{\varepsilon}_i^2$ for $i = 1, \ldots, n$. Here, $\mathsf{E}^*$ denotes the bootstrap expectation operator. The bootstrap responses $\{Y_i^*\}_{i=1}^n$ are defined by

$$Y_i^* = \langle \hat{\beta}_{g_n}, X_i \rangle + \varepsilon_i^*, \quad i = 1, \ldots, n,$$

where the estimator $\hat{\beta}_{g_n} \equiv \hat{\Gamma}_{g_n}^{-1} \hat{\Delta}_n$ plays the role of the true parameter $\beta$ in the bootstrap world. The bootstrap version $\hat{\beta}_{h_n}^*$ of the original data estimator $\hat{\beta}_{h_n}$ is finally defined based on the bootstrap data $\{(X_i, Y_i^*)\}_{i=1}^n$ with the same truncation level $h_n$.

For constructing bootstrap errors, we particularly consider multiplier wild bootstrap. Namely, the bootstrap errors $\{\varepsilon_i^*\}_{i=1}^n$ are defined as $\varepsilon_i^* \equiv W_i \hat{\varepsilon}_i$ for some iid copies $\{W_i\}_{i=1}^n$ of a random variable $W$ with $\mathsf{E}[W] = 0$ and $\mathsf{E}[W^2] = 1$ that are independent of the data sample $\{(X_i, Y_i)\}_{i=1}^n$. One of the most popular choices of the multipliers $\{W_i\}_{i=1}^n$ is the following two points distribution [3, 10, 11, 15, 16]:

$$\mathsf{P}\left(W_i = -\frac{\sqrt{5}-1}{2}\right) = \frac{\sqrt{5}+1}{2\sqrt{5}} = 1 - \mathsf{P}\left(W_i = \frac{\sqrt{5}+1}{2}\right).$$

As a continuous multiplier distribution, [10, 15] consider the multipliers defined as $W_i = V_i/2 + (V_i^2 - 1)/2$ where the variables $\{V_i\}_{i=1}^n$ are iid standard normal variables. These two types of multipliers ensures that $\mathsf{E}^*[(\varepsilon_i^*)^3] = \hat{\varepsilon}_i^3$ as well as $\mathsf{E}^*[\varepsilon_i^*] = 0$ and $\mathsf{E}^*[(\varepsilon_i^*)^2] = \hat{\varepsilon}_i^2$. We might simply consider standard normal variables as multipliers as well.

We list the technical conditions for bootstrap consistency below. We define the eigengaps $\{\delta_j\}_{j=1}^\infty$ as $\delta_1 \equiv \gamma_1 - \gamma_2$ and $\delta_j \equiv \min\{\gamma_j - \gamma_{j+1}, \gamma_{j-1} - \gamma_j\}$ for $j \geq 2$.

(A1) $\sup_{j\in\mathbb{N}} \gamma_j^{-2} \mathsf{E}[\langle X, \phi_j \rangle^4] < \infty$;

(A2) $\gamma_j$ is a convex function of $j \geq J$ for some integer $J \geq 1$;

(A3) $\sup_{j\in\mathbb{N}} \gamma_j j \log j < \infty$;

(A4) $n^{-1} \sum_{j=1}^{h_n} \delta_j^{-2} \to 0$ as $n \to \infty$;

(A5) $h_n s_{h_n}(X)^{-1} = O_{\mathsf{P}}(1)$;

(A6) $\sup_{j\in\mathbb{N}} \lambda_j^{-2}\mathsf{E}[\langle X\varepsilon, \psi_j\rangle^4] < \infty$, where $\lambda_j$ and $\psi_j$ are the $j$-th eigenvalue–eigenfunction pair of $\Lambda$;

(A7) $\sup_{j\in\mathbb{N}} \gamma_j^{-1}\|\Lambda^{1/2}\phi_j\|^2 < \infty$.

Conditions (A1) and (A6) ensure that $X$ and $X\varepsilon$ respectively have finite fourth moments. Conditions (A2)-(A4) are technical assumptions related to the decay behaviors of eigenvalues $\{\gamma_j\}$ and eigengaps $\{\delta_j\}$. Condition (A5) provides a mild lower bound for scaling $s_{h_n}(X_0)$. Condition (A7) is a technical condition that balances the eignedecay of $\Gamma$ and the decay rate of $\Lambda$ in terms of $\{\phi_j\}_{j=1}^{\infty}$. In addition to Conditions (A1)-(A7), the following condition is imposed to remove bias in the limit:

Condition $B(u)$ : $\sup_{j\in\mathbb{N}} j^{-1}m(j,u)\langle\beta,\phi_j\rangle^2 < \infty$,

depending on a generic constant $u > 0$ and function $m(j,u)$ of integer $j \geq 1$ defined as

$$m(j,u) = \max\left\{j^u, \sum_{l=1}^{j}\delta_l^{-2}\right\}. \tag{4.6}$$

We estimate the sampling distribution of $T_n(X_0)$ from (4.5) with the bootstrap distribution of

$$T_n^*(X_0) \equiv \sqrt{\frac{n}{s_{h_n}(X_0)}}[\langle\hat{\beta}_{h_n}^*, X_0\rangle - \langle\hat{\beta}_{g_n}, X_0\rangle]. \tag{4.7}$$

**Theorem 11.** *Suppose that Conditions (A1)-(A7) and (M) hold,* $k_n^{-1} + n^{-1/2}k_n^2\log k \to 0$, *and there exists* $\delta \in (0, 2]$ *such that* $\mathsf{E}[\|X\|^{4+2\delta}]$, $\mathsf{E}[|W|^{2+\delta}] < \infty$, *and* $n^{-\delta/2}h_n^{\delta/2}\sum_{j=1}^{h_n}\lambda_j^{-1-\delta/2} = O(1)$. *Along with Condition* $B(u)$ *for some* $u > 7$, *we further suppose that* $\tau \equiv \lim_{n\to\infty} h_n/g_n \geq 1$, $g_n^{-1} + n^{-1/2}h_n^{7/2}(\log h_n)^4 \to 0$, *and* $n = O(m(h_n, u))$. *Then, as* $n \to \infty$, *the wild bootstrap (WB) projection estimator* $T_n^*(X_0)$ *from (4.7) approximates the projection estimator* $T_n$ *from (4.5) in the sense that*

$$\sup_{y\in\mathbb{R}} |\mathsf{P}^*(T_n^*(X_0) \leq y|X_0) - \mathsf{P}(T_n(X_0) \leq y|X_0)| \xrightarrow{\mathsf{P}} 0.$$

*Proof.* This is proved by Propositions 32 and 33 based on decompositions (4.8) and (4.9) along with convergence results in [18] and [19]. □

## 4.4 Technical details

The proof of Theorem 11 is based on the following decompositions:

$$\hat{\beta}_{h_n} - \beta = \Gamma_{h_n}^{-1} U_n + (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) U_n + (\hat{\Pi}_{h_n} - \Pi_{h_n})\beta + \Pi_{h_n}\beta - \beta, \tag{4.8}$$

$$\hat{\beta}_{h_n}^* - \hat{\beta}_{g_n} = \Gamma_{h_n}^{-1} U_n^* + (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}) U_n^* + \hat{\Pi}_{h_n}\hat{\beta}_{g_n} - \hat{\beta}_{g_n}. \tag{4.9}$$

The convergences of bias terms in $T_n(X_0)$ and $T_n^*(X_0)$ related to $(\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1})U_n$, $(\hat{\Pi}_{h_n} - \Pi_{h_n})\beta$, $\Pi_{h_n}\beta - \beta$, and $\hat{\Pi}_{h_n}\hat{\beta}_{g_n} - \hat{\beta}_{g_n}$ are provided in the supplement (Section S1.3 and the results involving Equation (S15)) of [19]. The weak convergence of the term related to $\Gamma_{h_n}^{-1}U_n$ is given in Proposition 3 of [19] and its supplement (Section S1.3). In Section 4.4.1, we prove the bootstrap bias term related to $(\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1})U_n^*$ converges to zero by using the perturbation theory [6, 18, 19]. Then, Section 4.4.2 devotes to prove the weak convergence of the bootstrap variance term related to $\Gamma_{h_n}^{-1}U_n^*$ by verifying the Lyapunov condition.

### 4.4.1 Bias term

**Lemma 50.** *Write $\tilde{U}_n^* \equiv n^{-1}\sum_{i=1}^n X_i\varepsilon_i^*$. Then, we have*

$$\mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j} \|(zI - \Gamma)^{-1/2}\tilde{U}_n^*\|^2\right] \le Q_{1jn} + Q_{2jn}\|\hat{\beta}_{k_n} - \beta\|^2,$$

*where $Q_{1jn}$ and $Q_{2jn}$ are non-negative random variables such that*

$$\mathsf{E}[Q_{1jn}] \le \begin{cases} Cn^{-1}\delta_j^{-1} & \text{in general when } \mathsf{E}[\|X\varepsilon\|^2] < \infty \\ Cn^{-1} j\log j & \text{if either } \mathsf{E}[\varepsilon^2|X] = \sigma_\varepsilon^2 \in (0,\infty) \text{ or } \mathsf{E}[\varepsilon^4] < \infty, \end{cases}$$

*and $\mathsf{E}[Q_{2jn}] \le Cn^{-1} j \log j$.*

*Proof.* Note that

$$\begin{aligned} &\|(zI - \Gamma)^{-1/2}\tilde{U}_n^*\|^2 \\ =& n^{-2}\sum_{i=1}^n \|(zI-\Gamma)^{-1/2}X_i\|^2(\varepsilon_i^*)^2 \\ &+ n^{-2}\sum_{i=1}^n \langle (zI-\Gamma)^{-1/2}X_i, (zI-\Gamma)^{-1/2}X_{i'}\rangle \varepsilon_i^*\varepsilon_{i'}^*, \end{aligned}$$

which implies that

$$
\begin{aligned}
&\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\tilde{U}_n^*\|^2\\
\leq &n^{-2}\sum_{i=1}^n\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2(\varepsilon_i^*)^2\\
&+n^{-2}\sum_{i=1}^n\sup_{z\in\mathcal{B}_j}\langle(zI-\Gamma)^{-1/2}X_i,(zI-\Gamma)^{-1/2}X_{i'}\rangle\varepsilon_i^*\varepsilon_{i'}^*.
\end{aligned}
$$

Since $\mathsf{E}^*[(\varepsilon_i^*)^2]=\hat{\varepsilon}_{i,k_n}^2$ for each $i$ and $\mathsf{E}^*[\varepsilon_i^*\varepsilon_{i'}^*]=0$ if $i\neq i'$, we have

$$
\begin{aligned}
&\mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\tilde{U}_n^*\|^2\right]\leq n^{-2}\sum_{i=1}^n\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2\hat{\varepsilon}_{i,k_n}^2\\
\leq &2n^{-2}\sum_{i=1}^n\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2\varepsilon_i^2+2n^{-2}\sum_{i=1}^n\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2\langle\hat{\beta}_{k_n}-\beta,X_i\rangle^2\\
= &2n^{-2}\sum_{i=1}^n\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2\varepsilon_i^2+2\left\langle\left(n^{-2}\sum_{i=1}^n\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2X_i^{\otimes 2}\right)(\hat{\beta}_{k_n}-\beta),\hat{\beta}_{k_n}-\beta\right\rangle
\end{aligned}
\tag{4.10}
$$

The first term in (4.10) is differently bounded depending on the error assumption. We first consider general case with $\mathsf{E}[\|X\varepsilon\|^2]<\infty$. Recall by [13, Equation (5.3)], for $z\in\mathcal{B}_j$,

$$
\|(zI-\Gamma)^{-1/2}\|_\infty=\left(\min_{l\in\mathbb{N}}|z-\gamma_l|^{1/2}\right)^{-1}=|z-\gamma_j|^{-1/2}=(\delta_j/2)^{-1/2}.
$$

This implies that

$$
\begin{aligned}
\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^2\varepsilon^2\right]&\leq\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\|_\infty\|X\varepsilon\|^2\right]\\
&\leq 2\delta_j^{-1}\mathsf{E}[\|X\varepsilon\|^2]=2\mathrm{tr}(\mathsf{var}[X\varepsilon])\delta_j^{-1}.
\end{aligned}
$$

Next, under homoscedasticity with $\mathsf{E}[\varepsilon^2|X]=\sigma_\varepsilon^2$, we have that

$$
\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^2\varepsilon^2\right]=\sigma_\varepsilon^2\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^2\right]\leq Cj\log j.
$$

Last, under heteroscedasticity with $\mathsf{E}[\varepsilon^4]<\infty$, by Cauchy-Schwarz inequality, we have that

$$
\begin{aligned}
\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^2\varepsilon^2\right]&\leq\sqrt{\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^4\right]}\sqrt{\mathsf{E}[\varepsilon^4]}\\
&\leq Cj\log j.
\end{aligned}
$$

We thus have that

$$\mathsf{E}\left[n^{-2}\sum_{i=1}^{n}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2\varepsilon_i^2\right]$$
$$\leq\begin{cases}Cn^{-1}\delta_j^{-1} & \text{in general when } \mathsf{E}[|X\varepsilon\|^2]<\infty\\ Cn^{-1}j\log j & \text{if either } \mathsf{E}[\varepsilon^2|X]=\sigma_\varepsilon^2\in(0,\infty) \text{ or } \mathsf{E}[\varepsilon^4]<\infty.\end{cases}$$

The second term in (4.10) is bounded as

$$\begin{aligned}&\mathsf{E}\left[\left|n^{-2}\sum_{i=1}^{n}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2X_i^{\otimes 2}\right|\right]\\ \leq& n^{-1}\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2\|X_i\|^2\right]\\ \leq& n^{-1}\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^4\right]^{1/2}\mathsf{E}[\|X_i\|^4]^{1/2}\\ \leq& Cn^{-1}j\log j\end{aligned}$$

by the third part of Lemma 15. This completes the proof. □

**Lemma 51.** *Write* $U_n^*\equiv n^{-1}\sum_{i=1}^n(X_i-\bar{X})(\varepsilon_i^*-\bar{\varepsilon}^*)=n^{-1}\sum_{i=1}^n X_i\varepsilon_i^*-\bar{X}\bar{\varepsilon}^*$*. Then, we have that*

$$\mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}U_n^*\|^2\right]\leq Q_{1jn}+Q_{2jn}\|\hat{\beta}_{k_n}-\beta\|^2+R_nQ_{2jn},$$

*where* $Q_{1jn}$*,* $Q_{2jn}$*,* $Q_{3jn}$*, and* $R_n$ *are non-negative random variables such that*

$$\mathsf{E}[Q_{1jn}]\leq\begin{cases}Cn^{-1}\delta_j^{-1} & \text{in general when } \mathsf{E}[|X\varepsilon\|^2]<\infty\\ Cn^{-1}j\log j & \text{if either } \mathsf{E}[\varepsilon^2|X]=\sigma_\varepsilon^2\in(0,\infty) \text{ or } \mathsf{E}[\varepsilon^4]<\infty,\end{cases}$$

$\mathsf{E}[Q_{2jn}]\leq Cn^{-1}j\log j$*,* $\mathsf{E}[Q_{3jn}]\leq Cj\log j$*, and* $R_n=O_{\mathsf{P}}(n^{-1})$*.*

*Proof.* Note that

$$\|(zI-\Gamma)^{-1/2}U_n^*\|^2\leq 2\|(zI-\Gamma)^{-1/2}\tilde{U}_n^*\|^2+2(\bar{\varepsilon}^*)^2\|(zI-\Gamma)^{-1/2}\bar{X}\|^2. \tag{4.11}$$

The first term in (4.11) is bounded based on Lemma 50. For the second term in (4.11), note that

$$\mathsf{E}^*[(\bar{\varepsilon}^*)^2]=n^{-2}\sum_{i=1}^{n}\mathsf{E}^*[(\varepsilon_i^*)^2]+n^{-2}\sum_{i\neq i'}\mathsf{E}^*[\varepsilon_i^*\varepsilon_{i'}^*]=n^{-2}\sum_{i=1}^{n}\hat{\varepsilon}_{i,k_n}^2=O_{\mathsf{P}}(n^{-1})$$

if $\|\hat{\beta}_{k_n} - \beta\| \xrightarrow{\mathsf{P}} 0$, since

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n} \hat{\varepsilon}_{i,k_n}^2 &= n^{-1}\sum_{i=1}^{n}(\varepsilon_i - \langle X_i, \hat{\beta}_{k_n} - \beta\rangle)^2 \\
&\le n^{-1}\sum_{i=1}^{n}\varepsilon_i^2 + n^{-1}\sum_{i=1}^{n}\|X_i\|^2\|\hat{\beta}_{k_n} - \beta\|^2 + 2n^{-1}\sum_{i=1}^{n}\|X_i\varepsilon_i\|\|\hat{\beta}_{k_n} - \beta\| \\
&= \{\mathsf{E}[\varepsilon^2] + o_{\mathsf{P}}(1)\} + \{\mathsf{E}[\|X\|^2] + o_{\mathsf{P}}(1)\}o_{\mathsf{P}}(1) + \{\mathsf{E}[\|X\varepsilon\|] + o_{\mathsf{P}}(1)\}o_{\mathsf{P}}(1) \\
&= O_{\mathsf{P}}(1).
\end{aligned}
$$

Meanwhile, by Jensen's inequality, we have

$$
\begin{aligned}
&\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\bar{X}\|^2\right] \\
\le& n^{-1}\sum_{i=1}^{n}\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2\right] = \mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^2\right] \\
\le& Cj\log j.
\end{aligned}
$$

This completes the proof. □

**Proposition 32.** *Suppose Conditions (A1)-(A6) hold. As $n\to\infty$, we further suppose either*

1. $n^{-1/2}h_n^{-1/2}\left(\sum_{j=1}^{h_n}\delta_j^{-1}j\log j\right)^{1/2}\left\{\sum_{j=1}^{h_n}(j\log j)^2\right\}^{1/2}\to 0$ *in general when* $\mathsf{E}[\|X\varepsilon\|^2]<\infty$*; or*
2. $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}(j\log j)^2\to 0$ *if either* $\mathsf{E}[\varepsilon^2|X]\equiv\sigma_\varepsilon^2\in(0,\infty)$ *or* $\mathsf{E}[\varepsilon^4]<\infty$.

*Wh then have that*

$$
\mathsf{P}^*\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}|\langle(\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1})U_n^*, X_0\rangle|>\eta\Big|X_0\right)\xrightarrow{\mathsf{P}}0
$$

*for each $\eta>0$. Both convergence rates hold if $n^{-1/2}h_n^{7/2}(\log h_n)^4=O(1)$.*

*Proof.* We observe from Lemma 14 that

$$
\begin{aligned}
\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1} &= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}\left\{(zI-\hat{\Gamma}_n)^{-1}-(zI-\Gamma)^{-1}\right\}dz + r_{2n}\mathbb{I}_{\mathcal{A}_{h_n}^c} \\
&= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}(zI-\Gamma)^{-1/2}K_n(z)G_n(z)(zI-\Gamma)^{-1/2}dz + r_{2n}\mathbb{I}_{\mathcal{A}_{h_n}^c}.
\end{aligned}
$$

This implies that $|\langle(\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1})U_n^*, X_0\rangle| \leq C\sum_{j=1}^{h_n} A_j + \|r_{2n}\|_\infty \|U_n^*\|\|X_0\|\mathbb{I}_{\mathcal{A}_{h_n}^c}$, where

$$A_j = \int_{\mathcal{B}_j} \frac{1}{|z|}\|(zI-\Gamma)^{-1/2}X_0\|\|K_n(z)\|_\infty\|G_n(z)\|_\infty\|(zI-\Gamma)^{-1/2}U_n^*\|dz.$$

Note that

$$\begin{aligned}
&\mathsf{E}^*[A_j\mathbb{I}_{\mathcal{E}_j}]\\
\leq& C\mathrm{diam}(\mathcal{B}_j)\delta_j^{-1}\sup_{z\in\mathcal{B}_j}\|K_n(z)\|_\infty\mathbb{I}_{\mathcal{E}_j}\sup_{z\in\mathcal{B}_j}\|G_n(z)\|_\infty\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\|\mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}U_n^*\|\right]\\
\leq& CV_{1jn}V_{2jn}\mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}U_n^*\|^2\right]^{1/2},
\end{aligned}$$

where

$$V_{1jn}\equiv\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_0\| \quad\text{and}\quad V_{2jn}\equiv\sup_{z\in\mathcal{B}_j}\|K_n(z)\|_\infty\mathbb{I}_{\mathcal{E}_j}\sup_{z\in\mathcal{B}_j}\|G_n(z)\|_\infty.$$

Then,

$$\begin{aligned}
\mathsf{E}^*\left[\sum_{j=1}^{h_n}A_j\mathbb{I}_{\mathcal{E}_j}\right] &\leq C\sum_{j=1}^{h_n}V_{1jn}V_{2jn}\mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}U_n^*\|^2\right]^{1/2}\\
&\leq C\left(\sum_{j=1}^{h_n}V_{1jn}^2\right)^{1/2}\left(\sum_{j=1}^{h_n}V_{2jn}^2\mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}U_n^*\|^2\right]\right)^{1/2}.
\end{aligned}$$

By Lemma 15 and Lemma 16,

$$\mathsf{E}\left[\sum_{j=1}^{h_n}V_{1jn}^2\right]\leq C\sum_{j=1}^{h_n}(j\log j)^2.$$

Also, note by Lemma 51 and the independence between $\{(X_i,Y_i)\}_{i=1}^n$ and $X_0$ that

$$\mathsf{E}\left[\sum_{j=1}^{h_n}V_{2jn}^2Q_{1jn}\right]\leq\begin{cases}Cn^{-2}\sum_{j=1}^{h_n}\delta_j^{-1}(j\log j)^2 & \text{in general when } \mathsf{E}[|X\varepsilon\|^2]<\infty\\ Cn^{-2}\sum_{j=1}^{h_n}(j\log j)^3 & \text{if either } \mathsf{E}[\varepsilon^2|X]=\sigma_\varepsilon^2\in(0,\infty) \text{ or } \mathsf{E}[\varepsilon^4]<\infty,\end{cases}$$

$\mathsf{E}\left[\sum_{j=1}^{h_n} V_{2jn}^2 Q_{2jn}\right] \le Cn^{-2}\sum_{j=1}^{h_n}(j\log j)^3$, and $\mathsf{E}\left[\sum_{j=1}^{h_n} V_{2jn}^2 Q_{3jn}\right] \le C\sum_{j=1}^{h_n}(j\log j)^3$. This implies that

$$\begin{aligned}
&\sum_{j=1}^{h_n} V_{2jn}^2 \mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}U_n^*\|^2\right]\\
\le&\sum_{j=1}^{h_n} V_{2jn}^2 Q_{1jn} + \sum_{j=1}^{h_n} V_{2jn}^2 Q_{2jn}\|\hat{\beta}_{k_n}-\beta\|^2 + \sum_{j=1}^{h_n} V_{2jn}^2 Q_{3jn}R_n\\
=&\begin{cases} O_{\mathsf{P}}\left(n^{-2}\sum_{j=1}^{h_n}\delta_j^{-1}(j\log j)^2\right) & \text{in general when } \mathsf{E}[\|X\varepsilon\|^2]<\infty\\ O_{\mathsf{P}}\left(n^{-2}\sum_{j=1}^{h_n}(j\log j)^3\right) & \text{if either } \mathsf{E}[\varepsilon^2|X]=\sigma_\varepsilon^2\in(0,\infty) \text{ or } \mathsf{E}[\varepsilon^4]<\infty,\end{cases}
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathsf{E}^*\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}\sum_{j=1}^{h_n}A_j\mathbb{I}_{\mathcal{E}_j}\right]\\
=&\begin{cases} O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\left\{\sum_{j=1}^{h_n}\delta_j^{-1}(j\log j)^2\right\}^{1/2}\left\{\sum_{j=1}^{h_n}(j\log j)^2\right\}^{1/2}\right)\\ \text{in general when } \mathsf{E}[\|X\varepsilon\|^2]<\infty\\ O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\left\{\sum_{j=1}^{h_n}(j\log j)^3\right\}^{1/2}\left\{\sum_{j=1}^{h_n}(j\log j)^2\right\}^{1/2}\right)\\ \text{if either } \mathsf{E}[\varepsilon^2|X]=\sigma_\varepsilon^2\in(0,\infty) \text{ or } \mathsf{E}[\varepsilon^4]<\infty,\end{cases}
\end{aligned}$$

and by following the argument of Remark 6, we have the desired result. Note that
$n^{-1/4}h_n^{1/4}\left(\sum_{j=1}^{h_n}\delta_j^{-1}\right)^{1/2} \le \left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}\right)^{1/4} \to 0$ by Cauchy-Schwarz inequality. It then follows that

$$\begin{aligned}
&n^{-1/2}h_n^{-1/2}\left\{\sum_{j=1}^{h_n}\delta_j^{-1}(j\log j)^2\right\}^{1/2}\left\{\sum_{j=1}^{h_n}(j\log j)^2\right\}^{1/2}\\
\le&n^{-1/2}h_n^2(\log h_n)^2\left(\sum_{j=1}^{h_n}\delta_j^{-1}\right)^{1/2} = n^{-1/4}h_n^{7/4}(\log h_n)^2 n^{-1/4}h_n^{-1/4}\left(\sum_{j=1}^{h_n}\delta_j^{-1}\right)^{1/2}\\
=&o(\{n^{-1/2}h_n^{7/2}(\log h_n)^4\}^{1/2})
\end{aligned}$$

and

$$n^{-1/2}h_n^{-1/2}\left\{\sum_{j=1}^{h_n}(j\log j)^3\right\}^{1/2}\left\{\sum_{j=1}^{h_n}(j\log j)^2\right\}^{1/2} \le n^{-1/2}h_n^3(\log h_n)^{5/2}$$

$$\le o(n^{-1/2}h_n^{7/2}(\log h_n)^4).$$

This proves the last assertion. □

### 4.4.2 Weak convergence of variance term

**Lemma 52.** *As $n \to \infty$, if $\|\hat{\beta}_{k_n} - \beta\| \xrightarrow{\mathsf{P}} 0$, we have that $\mathsf{E}^*[\|\bar{X}\bar{\varepsilon}^*\|^2] = O_{\mathsf{P}}(n^{-2})$, where $\bar{X} \equiv n^{-1}\sum_{i=1}^n X_i$ and $\bar{\varepsilon}^* \equiv n^{-1}\sum_{i=1}^n \varepsilon_i^*$.*

*Proof.* Note that

$$(\bar{\varepsilon}^*)^2 = n^{-2}\sum_{i=1}^n(\varepsilon_i^*)^2 + n^{-2}\sum_{i\neq i'}\varepsilon_i^*\varepsilon_{i'}^*,$$

which implies that $\mathsf{E}^*[(\bar{\varepsilon}^*)^2] = n^{-2}\sum_{i=1}^n \hat{\varepsilon}_{i,k_n}^2$. Since

$$\begin{aligned}
n^{-1}\sum_{i=1}^n \hat{\varepsilon}_{i,k_n}^2 &= n^{-1}\sum_{i=1}^n(\varepsilon_i - \langle X_i, \hat{\beta}_{k_n} - \beta\rangle)^2 \\
&\le n^{-1}\sum_{i=1}^n \varepsilon_i^2 + n^{-1}\sum_{i=1}^n \|X_i\|^2\|\hat{\beta}_{k_n} - \beta\|^2 + 2n^{-1}\sum_{i=1}^n \|X_i\varepsilon_i\|\|\hat{\beta}_{k_n} - \beta\| \\
&= \{\mathsf{E}[\varepsilon^2] + o_{\mathsf{P}}(1)\} + \{\mathsf{E}[\|X\|^2] + o_{\mathsf{P}}(1)\}o_{\mathsf{P}}(1) + \{\mathsf{E}[\|X\varepsilon\|] + o_{\mathsf{P}}(1)\}o_{\mathsf{P}}(1) \\
&= O_{\mathsf{P}}(1),
\end{aligned}$$

we have that $\mathsf{E}^*[(\bar{\varepsilon}^*)^2] = O_{\mathsf{P}}(n^{-1})$. Finally, since $\bar{X} = O_{\mathsf{P}}(n^{-1/2})$ [cf. 12, Theorem 2.3], we conclude that $\mathsf{E}^*[\|\bar{X}\bar{\varepsilon}^*\|^2] = \|\bar{X}\|^2\mathsf{E}^*[(\bar{\varepsilon}^*)^2] = O_{\mathsf{P}}(n^{-2})$. □

**Lemma 53.** *As $n \to \infty$, we have*

$$\mathsf{E}^*\left[\frac{n}{s_{h_n}(X_0)}\langle\bar{X}\bar{\varepsilon}^*, \Gamma_{h_n}^{-1}X_0\rangle^2\middle| X_0\right] = o_{\mathsf{P}}(1),$$

*where $\bar{X} \equiv n^{-1}\sum_{i=1}^n X_i$ and $\bar{\varepsilon}^* \equiv n^{-1}\sum_{i=1}^n \varepsilon_i^*$.*

*Proof.* Note from Jensen's inequality that

$$n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1} \le n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\gamma_j^{-1} \le \left(n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-2}\right)^{1/2} \le \left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}\right)^{1/2}$$

From Lemma 52, since $\mathsf{E}[\|\Gamma_{h_n}^{-1}X_0\|^2] = \sum_{j=1}^{h_n}\gamma_j^{-1}$, we have that

$$\begin{aligned}
&\mathsf{E}^*\left[\frac{n}{s_{h_n}(X_0)}\langle \bar{X}\bar{\varepsilon}^*, \Gamma_{h_n}^{-1}X_0\rangle^2 \Big| X_0\right] \\
\le&\{h_n s_{h_n}(X_0)^{-1}\}(n\mathsf{E}^*[\|\bar{X}\bar{\varepsilon}^*\|^2])(h_n^{-1}\|\Gamma_{h_n}^{-1}X_0\|^2) \\
=&O_{\mathsf{P}}(1)O_{\mathsf{P}}(n^{-1})O_{\mathsf{P}}\left(h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right) \\
=&O_{\mathsf{P}}\left(n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right) = o_{\mathsf{P}}(n^{-1/2})
\end{aligned}$$

by Condition (A5). □

**Proposition 33.** *Suppose that $n^{-\delta/2}h_n^{\delta/2}\sum_{j=1}^{h_n}\lambda_j^{-(2+\delta)/2} = O(1)$ and $\mathsf{E}[\|X\|^{4+2\delta}] < \infty$ hold for some $\delta \in (0,2]$, $\|\hat{\beta}_{k_n} - \beta\| \xrightarrow{\mathsf{P}} 0$ as $n \to \infty$, and Condition (A7) hold. Then, as $n \to \infty$, if $n^{-\delta/2}h_n^{(2+\delta)/2} \to 0$, we have that*

$$\sup_{y\in\mathbb{R}}\left|\mathsf{P}^*\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle\Gamma_{h_n}^{-1}U_n^*, X_0\rangle \le y \Big| X_0\right) - \Phi(y)\right| \xrightarrow{\mathsf{P}} 0,$$

*where $\Phi$ denotes the cumulative distribution function of the standard normal distribution.*

*Proof.* The bootstrap variance term is $v_n^* \equiv \sqrt{n/s_{h_n}(X_0)}\langle\Gamma_{h_n}^{-1}U_n^*, X_0\rangle$ where $U_n^* \equiv n^{-1}\sum_{i=1}^n(X_i - \bar{X})(\varepsilon_i^* - \bar{\varepsilon}^*) = n^{-1}\sum_{i=1}^n X_i\varepsilon_i^* - \bar{X}\bar{\varepsilon}^*$. We know that the latter term related to $\bar{X}\bar{\varepsilon}^*$ is negligible due to Lemma 53. Write $Z_{i,n}^* \equiv \langle X_i\varepsilon_i^*, \Gamma_{h_n}^{-1}X_0\rangle$ so that $\mathsf{E}^*[Z_{i,n}^*|X_0] = 0$ and $\mathsf{E}^*[(Z_{i,n}^*)^2|X_0] = \langle X_i\hat{\varepsilon}_{i,k_n}, \Gamma_{h_n}^{-1}X_0\rangle^2$.

We first derive that

$$n^{-1}\hat{v}_n^2 \sim_{\mathsf{P}} s_{h_n}(X_0), \tag{4.12}$$

in the sense that

$$\frac{n^{-1}\hat{v}_n^2}{s_{h_n}(X_0)} \xrightarrow{\mathsf{P}} 1.$$

Note that

$$
\begin{aligned}
n^{-1}\hat{v}_n^2 \equiv n^{-1}\sum_{i=1}^{n}\mathsf{E}^*[(Z_{i,n}^*)^2|X_0] &= n^{-1}\sum_{i=1}^{n}\langle X_i\hat{\varepsilon}_{i,k_n},\Gamma_{h_n}^{-1}X_0\rangle^2 \\
&= n^{-1}\sum_{i=1}^{n}\langle (X_i\hat{\varepsilon}_{i,k_n})^{\otimes 2}\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\rangle \\
&= \left\langle \left(n^{-1}\sum_{i=1}^{n}(X_i\hat{\varepsilon}_{i,k_n})^{\otimes 2}-\Lambda\right)\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\right\rangle + s_{h_n}(X_0).
\end{aligned}
$$

Thus, Proposition 20 proves (4.12).

Write $\mathcal{L}_n = \mathcal{L}_{n,\delta} \equiv \hat{v}_n^{-(2+\delta)}\sum_{i=1}^{n}\mathsf{E}^*[|Z_{i,n}^*|^{2+\delta}|X_0]$ for the Lyapunov term. We will show that

$$
\mathcal{L}_n \xrightarrow{\mathsf{P}} 0 \tag{4.13}
$$

Note that

$$
\begin{aligned}
\mathcal{L}_n = \hat{v}_n^{-(2+\delta)}\sum_{i=1}^{n}\mathsf{E}^*[|Z_{i,n}^*|^{2+\delta}|X_0] &\le \hat{v}_n^{-(2+\delta)}\sum_{i=1}^{n}\mathsf{E}^*[\|\Lambda_{h_n}^{-1/2}X_i\varepsilon_i^*\|^{2+\delta}|X_0]\|\Lambda_{h_n}^{1/2}\Gamma_{h_n}^{-1}X_0\|^{2+\delta} \\
= \left(\frac{\|\Lambda_{h_n}^{1/2}\Gamma_{h_n}^{-1}X_0\|^2}{n^{-1}\hat{v}_n^2}\right)^{(2+\delta)/2} & n^{-\delta/2}\mathsf{E}^*\left[n^{-1}\sum_{i=1}^{n}\|\Lambda_{h_n}^{-1/2}X_i\varepsilon_i^*\|^{2+\delta}\Big|X_0\right].
\end{aligned}
$$

Since

$$
\frac{\|\Lambda_{h_n}^{1/2}\Gamma_{h_n}^{-1}X_0\|^2}{n^{-1}\hat{v}_n^2} \le \frac{\|\Lambda^{1/2}\Gamma_{h_n}^{-1}X_0\|^2}{n^{-1}\hat{v}_n^2} = \frac{s_{h_n}(X_0)}{n^{-1}\hat{v}_n^2} \xrightarrow{\mathsf{P}} 1,
$$

we have

$$
\frac{\|\Lambda_{h_n}^{1/2}\Gamma_{h_n}^{-1}X_0\|^2}{n^{-1}\hat{v}_n^2} = O_{\mathsf{P}}(1).
$$

Note by Lyapunov inequality that

$$
\begin{aligned}
\left(h_n^{-1}\sum_{j=1}^{h_n}a_j^2\right)^{1/2} &\le \left(h_n^{-1}\sum_{j=1}^{h_n}a_j^{2+\delta}\right)^{1/(2+\delta)} \\
\iff \left(h_n^{-1}\sum_{j=1}^{h_n}a_j^2\right)^{(2+\delta)/2} &\le h_n^{-1}\sum_{j=1}^{h_n}a_j^{2+\delta} \\
\iff \left(\sum_{j=1}^{h_n}a_j^2\right)^{(2+\delta)/2} &\le h_n^{\delta/2}\sum_{j=1}^{h_n}a_j^{2+\delta}
\end{aligned}
$$

We then have

$$
\begin{aligned}
&\mathsf{E}^*[\|\Lambda_{h_n}^{-1/2}X_i\varepsilon_i^*\|^{2+\delta}] = \|\Lambda_{h_n}^{-1/2}X_i\hat{\varepsilon}_{i,k_n}\|^{2+\delta}\mathsf{E}[|W|^{2+\delta}]\\
=&\mathsf{E}[|W|^{2+\delta}]\left(\sum_{j=1}^{h_n}\lambda_j^{-1}\langle X\hat{\varepsilon}_{i,k_n},\psi_j\rangle^2\right)^{(2+\delta)/2} \leq \mathsf{E}[|W|^{2+\delta}]h_n^{\delta/2}\sum_{j=1}\lambda_j^{-(2+\delta)/2}|\langle X_i\hat{\varepsilon}_{i,k_n},\psi_j\rangle|^{2+\delta}\\
\leq&\mathsf{E}[|W|^{2+\delta}]2^{1+\delta}h_n^{\delta/2}\sum_{j=1}^{h_n}\lambda_j^{-(2+\delta)/2}|\langle X_i(\hat{\varepsilon}_{i,k_n}-\varepsilon_i),\psi_j\rangle|^{2+\delta}\\
&+\mathsf{E}[|W|^{2+\delta}]2^{1+\delta}h_n^{\delta/2}\sum_{j=1}^{h_n}\lambda_j^{-(2+\delta)/2}|\langle X_i\varepsilon_i,\psi_j\rangle|^{2+\delta} \qquad (4.14)
\end{aligned}
$$

from Jensen's inequality with form

$$
\left(\frac{x+y}{2}\right)^a \leq \frac{x^a+y^a}{2}, \quad \forall a>0, x,y\geq 0.
$$

Since $\hat{\varepsilon}_{i,k_n}-\varepsilon_i = -\langle\hat{\beta}_{k_n}-\beta, X_i\rangle$, the first term in (4.14) is

$$
\begin{aligned}
&h_n^{\delta/2}\sum_{j=1}^{h_n}\lambda_j^{-(2+\delta)/2}|\langle X_i(\hat{\varepsilon}_{i,k_n}-\varepsilon_i),\psi_j\rangle|^{2+\delta}\\
=&h_n^{\delta/2}\sum_{j=1}^{h_n}\lambda_j^{-(2+\delta)/2}|\langle X_i^{\otimes 2}(\hat{\beta}_{k_n}-\beta),\psi_j\rangle|^{2+\delta} = h_n^{\delta/2}\sum_{j=1}^{h_n}\lambda_j^{-(2+\delta)/2}|\langle\hat{\beta}_{k_n}-\beta, X_i^{\otimes 2}\psi_j\rangle|^{2+\delta}\\
\leq&h_n^{\delta/2}\sum_{j=1}^{h_n}\lambda_j^{-(2+\delta)/2}\|\hat{\beta}_{k_n}-\beta\|^{2+\delta}\|X_i\|^{4+2\delta}.
\end{aligned}
$$

Again, for $\delta\in(0,2]$, note by Lyapunov inequality that

$$
\begin{aligned}
&\left(h_n^{-1}\sum_{j=1}^{h_n}a_j^{2+\delta}\right)^{1/(2+\delta)} \leq \left(h_n^{-1}\sum_{j=1}^{h_n}a_j^4\right)^{1/4}\\
\Longleftrightarrow\ & h_n^{-1}\sum_{j=1}^{h_n}a_j^{2+\delta} \leq \left(h_n^{-1}\sum_{j=1}^{h_n}a_j^4\right)^{(2+\delta)/4}\\
\Longleftrightarrow\ & h_n^{\delta/2}\sum_{j=1}^{h_n}a_j^{2+\delta} \leq h_n^{(2+\delta)/4}\left(\sum_{j=1}^{h_n}a_j^4\right)^{(2+\delta)/4}.
\end{aligned}
$$

The second term in (4.14) is then

$$
h_n^{\delta/2}\sum_{j=1}\lambda_j^{-(2+\delta)/2}|\langle X_i\varepsilon_i,\psi_j\rangle|^{2+\delta} \leq h_n^{(2+\delta)/4}\left(\sum_{j=1}^{h_n}\lambda_j^{-2}\langle X_i\varepsilon_i,\psi_j\rangle^4\right)^{(2+\delta)/4}.
$$

Thus,

$$
\begin{aligned}
&n^{-\delta/2}\mathsf{E}^*\left[n^{-1}\sum_{i=1}^{n}\|\Lambda_{h_n}^{-1/2}X_i\varepsilon_i^*\|^{2+\delta}\Big|X_0\right]\\
\le&\mathsf{E}[|W|^{2+\delta}]2^{1+\delta}\left(n^{-\delta/2}h_n^{\delta/2}\sum_{j=1}^{h_n}\lambda_j^{-(2+\delta)/2}\right)\left(n^{-1}\sum_{i=1}^{n}\|X_i\|^{4+2\delta}\right)\|\hat{\beta}_{k_n}-\beta\|^{2+\delta} \qquad (4.15)\\
&+\mathsf{E}[|W|^{2+\delta}]2^{1+\delta}n^{-\delta/2}h_n^{(2+\delta)/4}n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{h_n}\lambda_j^{-2}\langle X_i\varepsilon_i,\psi_j\rangle^4\right)^{(2+\delta)/4}.
\end{aligned}
$$

The first term in (4.15) converges to zero in probability since $n^{-\delta/2}h_n^{\delta/2}\sum_{j=1}^{h_n}\lambda_j^{-(2+\delta)/2}=O(1)$, $\mathsf{E}[\|X\|^{4+2\delta}]<\infty$, and $\|\hat{\beta}_{k_n}-\beta\|\xrightarrow{\mathsf{P}}0$. In addition, since $\sup_{j\in\mathbb{N}}\lambda_j^{-2}\mathsf{E}[\langle X\varepsilon,\psi_j\rangle^4]<\infty$, by Lyapunov inequality,

$$
\mathsf{E}\left[\left(\sum_{j=1}^{h_n}\lambda_j^{-2}\langle X_i\varepsilon_i,\psi_j\rangle^4\right)^{(2+\delta)/4}\right]\le\left(\mathsf{E}\left[\sum_{j=1}^{h_n}\lambda_j^{-2}\langle X_i\varepsilon_i,\psi_j\rangle^4\right]\right)^{(2+\delta)/4}\le Ch_n^{(2+\delta)/4},
$$

and hence, the second term in (4.15) is bounded as

$$
\mathsf{E}\left[n^{-\delta/2}h_n^{(2+\delta)/4}n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{h_n}\lambda_j^{-2}\langle X_i\varepsilon_i,\psi_j\rangle^4\right)^{(2+\delta)/4}\right]\le Cn^{-\delta/2}h_n^{(2+\delta)/2}.
$$

Therefore, as $n\to\infty$, since $n^{-\delta/2}h_n^{(2+\delta)/2}\to 0$, then the second term converges to zero in probability, which verifies (4.13).

Finally, by combining Slutsky's theorem, Polya's theorem [1, Theorem 9.1.4], a subsequence argument [2, Theorem 20.5], Lemma 53, and (4.12), (4.13), we conclude the desired result. □

## 4.5 References


[1] Athreya, K. B., and Lahiri, S. N. *Measure theory and probability theory.* Springer Science & Business Media, 2006.

[2] Billingsley, P. *Probability and Measure.* John Wiley & Sons, 1999.

[3] Cao-Abad, R. Rate of convergence for the wild bootstrap in nonparametric regression. *The Annals of Statistics 19*, 4 (1991), 2226–2231.

[4] Cardot, H., Ferraty, F., and Sarda, P. Functional linear model. *Statistics & Probability Letters 45*, 1 (1999), 11–22.

[5] Cardot, H., Ferraty, F., and Sarda, P. Spline estimators for the functional linear model. *Statistica Sinica* (2003), 571–591.

[6] Cardot, H., Mas, A., and Sarda, P. Clt in functional linear regression models. *Probability Theory and Related Fields 138*, 3-4 (2007), 325–361.

[7] Dezeure, R., Bühlmann, P., and Zhang, C.-H. High-dimensional simultaneous inference with the bootstrap. *Test 26*, 4 (2017), 685–719.

[8] Ferraty, F., Van Keilegom, I., and Vieu, P. On the validity of the bootstrap in non-parametric functional regression. *Scandinavian Journal of Statistics 37*, 2 (2010), 286–306.

[9] González-Manteiga, W., and Martínez-Calvo, A. Bootstrap in functional linear regression. *Journal of Statistical Planning and Inference 141*, 1 (2011), 453–461.

[10] Härdle, W., and Mammen, E. Bootstrap methods in nonparametric regression. In *Nonparametric functional estimation and related topics*. Springer, 1991, pp. 111–123.

[11] Härdle, W., and Marron, J. S. Bootstrap simultaneous error bars for nonparametric regression. *The Annals of Statistics* (1991), 778–796.

[12] Horváth, L., and Kokoszka, P. *Inference for functional data with applications*, vol. 200. Springer Science & Business Media, 2012.

[13] Hsing, T., and Eubank, R. *Theoretical foundations of functional data analysis, with an introduction to linear operators*, vol. 997. John Wiley & Sons, 2015.

[14] Liu, R. Y. Bootstrap procedures under some non-iid models. *The annals of statistics 16*, 4 (1988), 1696–1708.

[15] Mammen, E. Bootstrap and wild bootstrap for high dimensional linear models. *The annals of statistics 21*, 1 (1993), 255–285.

[16] Mammen, E. Resampling methods for nonparametric regression. *Smoothing and Regression: Approaches, Computation, and Application* (2000), 425–450.

[17] Wu, C.-F. J. Jackknife, bootstrap and other resampling methods in regression analysis. *the Annals of Statistics 14*, 4 (1986), 1261–1295.

[18] Yeon, H., Dai, X., and Nordman, D. J. Bootstrap inference in functional linear regression models with scalar response. *accepted by Bernoulli Journal* (2022).

[19] Yeon, H., Dai, X., and Nordman, D. J. Bootstrap inference in functional linear regression models with scalar response under heteroscedasticity.

# CHAPTER 5. GENERAL CONCLUSION

This dissertation presents different bootstrap methods for inference in functional linear regression models (FLRMs). Central limit theorems for projection are studied as well, which are fundamental results themselves and are basis to verify bootstrap validity.

In Chapter 2, we developed a residual bootstrap in FLRMs with scalar responses and random functional regressors $X$. At a new target regressor $X_0$, inference about projections $\langle \beta, X_0 \rangle$ or future responses $Y_0$ is often useful, but it is often complicated to obtain the sampling distributions of regression estimators in FLRMs due to bias issues. We established theory to show that the bootstrap captures these sampling distributions, even conditionally on the observed set of data regressors $\{X_i\}_{i=1}^n$. In the bootstrap framework, the target regressor $X_0$ may be treated as given or unobserved, and bootstrap inference also extends to simultaneous inference at a collection $\mathcal{X}_0 \equiv \{X_{0,l}\}_{l=1}^m$ of regressors. In contrast, the simultaneous inference through normal theory-based approaches alone is often intractable in practice. Numerical studies also showed that the bootstrap outperforms intervals based on normal approximations when the latter applies. We also provided a rule of thumb for choosing the tuning parameters involved in the bootstrap for FLRMs, which was shown to exhibit good performance in simulations and was applied to a real data illustration. In developing the bootstrap, we refined a foundational central limit theorem for estimating projections $\langle \beta, X_0 \rangle$ in FLRMs.

In Chapter 3, we have developed a paired bootstrap (PB) for inference in FLRMs with general heteroscedastic errors. As a preliminary result, a central limit theorem under heteroscedasticity was established for estimated projections $\langle \beta, X_0 \rangle$ of the slope function $\beta$ onto a new predictor $X_0$, along with appropriate scaling $s_{h_n}(X_0)$ for self-normalization. Further, the projection of a functional principal component estimator $\hat{\beta}_{h_n}$ onto a new predictor $X_0$ can be successfully approximated by the PB for improved inference in finite samples. As such estimators $\hat{\beta}_{h_n}$ in

FLRMs involve truncation parameters $h_n$, a modified PB estimator was proposed to allow valid distributional approximations with the greatest possible flexibility in such truncation parameters for bootstrap. In contrast, a naive implementation of PB (i.e., adapted directly from standard multiple regression) can be shown to have less validity in application and becomes viable only for much narrower configuration of truncation parameters. The PB approach was also adapted to formulate new tests for assessing the orthogonality of the slope function to a subspace spanned by pre-selected regressor curves. Numerical studies showed that the existing residual bootstrap can fail under heterocedasticity, while PB can perform well in this context for interval estimation as well as for testing. We suggested also a PB implementation based on bootstrap studentization steps and provided a rule of thumb for selecting the two main tuning parameters (truncaton levels) involved in the PB.

In Chapter 4, we developed a wild bootstrap (WB) as an alternative of PB in Chapter 3. The proposed WB has computational advantages since it avoids repetition of spectral decompositions in PB procedure. The bootstrap consistency in mild assumptions are provided. As future work, the consistency of a bootstrap scaling should be verified to get valid inference when using a studentized statistic. The confidence interval from WB should also be numerically demonstrated by simulation studies; we expect that there are some large sample cases where WB might outperform PB though this may not always be the case.

Potential extensions of interest might include bootstrap methods in other functional linear models such as FLRMs with functional response or generalized functional linear models.